ACTIVITY PATTERNS IN A PARKINSON'S MONKEY MODEL

by
Andrew Leroux

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
April 2015

Abstract

In a study conducted at the University of Pittsburgh, 35 female rhesus monkeys were exposed to the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) in an effort to establish a plausible animal model for Parkinson's disease progression and treatment using neurotrophic factors. In addition to standard clinical measurements of Parkinson's severity, actigraphy devices recording minute level activity were fit to each of the monkeys. This thesis primarily aims to understand how activity patterns in healthy subjects can predict their susceptibility to developing Parkinson's-like symptoms as a result of MPTP exposure. Secondarily, we show how subjects respond to neurotrophic factors in terms of both subjects' activity and through standard clinical metrics.

Primary Reader: Ciprian Crainiceanu, PhD

Secondary Reader: Elizabeth Colantuoni, PhD

Acknowledgements

This document is the result of the better part of two years of study and work at the Johns Hopkins Bloomberg School of Public Health. Here, working toward a Masters degree in biostatistics, I have grown both as a person and as a professional statistician. I would like to take this space to thank those without whom this thesis would not have been possible. First, my thesis adviser Ciprian Crainiceanu has consistently been a wonderful mentor and has provided me with opportunities for which I will be forever grateful. In addition to Ciprian, Elizabeth Colantuouni, Haochang Shou, Luo Xiao, and Jonathan Gellar have all played a considerable role in helping to shape the way I approach statistical problems. I'd like to further thank all of the fantastic and brilliant faculty who taught me along the way. Also, this thesis would certainly not have been possible without the work and permission of Judy Cameron and everyone at the University of Pittsburgh who collected the data. Last (but never least), I would like to thank my parents and my partner Tushar who have always been my number one source of support and have never stopped pushing me to be the best I can be.

Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

Parkinson's disease is a chronic, progressive and ultimately debilitating neurological disease. Currently, treatments are limited to slowing disease progression and reducing the physical/neurological symptoms of the disease (Dutta et. al, 2013). The disease progresses as a result of the death of dopinergic neurons (Kong et. al 2015). Establishing which populations are at highest risk for developing Parkinson's disease (i.e. at risk populations) is an area of ongoing research. It is believed that exposure to environmental toxins play a key role in risk for developing Parkinson's disease. Specifically, the neurotoxin MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) has been shown to cause the onset of permanent Parkinson's symptoms in humans. The suspected cause for the onset of Parkinson's symptoms is the destruction of dopaminergic neurons resulting from the exposure (injection) of MPTP (Burns, 1983; Chassiah et. al, 2001).

Given that the destruction of the dopaminergic neurons is believed to be responsible for the onset of Parkinson's symptoms, neurotrophic factors that protect the dopaminergic neurons have potential as therapeutic treatments for slowing the process of neurological degradation and thus disease progression in Parkinson's patients (Yue et. al, 2013). One measurement that is highly correlated with severity of Parkinson's symptoms is overall activity (lower activity is highly associated with more severe disease symptoms) (Gomez-Mancilla et. al, 1993). To assess the protective benefits of several neurotrophic factors (NTF), a study at the University of Pittsburgh was conducted involving the sequential administration of MPTP and NTF in rhesus monkeys.

1.2 Data

This particular study involved 35 female rhesus monkeys observed over three distinct time periods: at baseline (2 weeks), 3-6 weeks after MPTP administration and 1-6 weeks after neurotrophic factor administration. For the purposes of this manuscript, we have assigned these monkeys arbitrary identification (ID) numbers 1-35. Throughout each of these phases, the monkeys wore accelerometers, providing minute level actigraphy data (1440 measurements per day per monkey). In addition to the activity measurements, the data contains Parkinson's severity scores for (most of) the monkeys. This severity score was assessed both after MPTP and after three rounds of neurotrophic factor administration. Hereafter, the severity score will be referred to as rating scale or rating score.

Not all 35 monkeys participated in the NTF phase of the study. Two monkeys were excluded as a result of an overly strong strong negative response to MPTP (one of which required additional therapies to stay alive following MPTP administration). Three monkeys were excluded due to a lack of clinical response to MPTP as measured by the rating scale. As a result, it was determined that NTF administration was inappropriate for these monkeys. Those monkeys that participated in the NTF portion of the experiment were given three infusions of NTF (or a placebo surgery). Each infusion was a major surgical procedure, and as such there are expected recovery times for activity following each infusion.

A number of days of activity data were excluded from the final activity data set. The reason for exclusion of days was usually related to data quality. That is, the device was found to have been malfunctioning or not working on a particular day/set of days. With the exception of malfunctions, monkeys wore the same accelerometer in each phase of the study.

Monkeys chosen for inclusion in the study were all approximately 17 years of age (life

span $\approx$ 25 years) at the start of the study. Additionally, all monkeys chosen exhibited no obvious symptoms of Parkinson's disease at entry.

1.3 Objectives

The main objectives of this analysis are to answer the following questions:

- Response to MPTP (Onset of Disease Symptoms):
  - Is there an observed response to MPTP (reflected in change in activity patterns) in rhesus monkeys?
  - Does average activity level at baseline predict response to MPTP?
  - Are there any other activity measures that can predict response to MPTP?
- Neurotrophic Factor (NTF) Infusion Effect (Recovery):
  - Does the administration of any NTF result in an improvement in rating scale?
  - Does the administration of any NTF result in an increase in average activity?

The analysis that addresses each of the objectives mentioned above (and the remainder of this document) is organized as follows: Exploratory data analysis (Chapter 2), analysis of monkey response to MPTP (Chapter 3), analysis of monkey response to NTF (Chapter 4), conclusions (Chapter 5)

2. Data Structure/EDA

2.1 Activity Counts

As mentioned previously, the device used in this study provides a single 'activity count' for every minute of the day. This minute level count is determined using an algorithm that summarizes the raw tri-axial acceleration data over a sixty second window. One concern when using these types of devices is that of positioning. That is, different locations on the body will be subject to different magnitudes of acceleration for the same type of activity. This concern is especially relevant in wrist-worn accelerometer studies (Bai et. al, 2014). In this study, monkeys had the devices placed around their necks and were reasonably secure in their placement (not much room to rotate around the neck). This placement is reasonably predicted to reduce variation in the magnitude of acceleration for the same activity and allow for more comparable activity summaries across monkeys.

A further crucial (and often overlooked) assumption in comparing activity numbers between monkeys is that activity counts across all devices are comparable. That is, will the same type of movement will result in exactly the same number of activity counts for any two randomly selected devices (Stephen & Spiro, 2001)? Although testing this assumption is beyond the scope of this analysis, the study used the same device within monkeys to the extent possible. As a result, the assumption that activity counts are consistent *at least* within individual monkeys seems reasonable. Furthermore, considering all devices were of the same make/manufacturer, activity counts should also be comparable between monkeys.

Structurally, one feature common to human and monkey activity count data is the

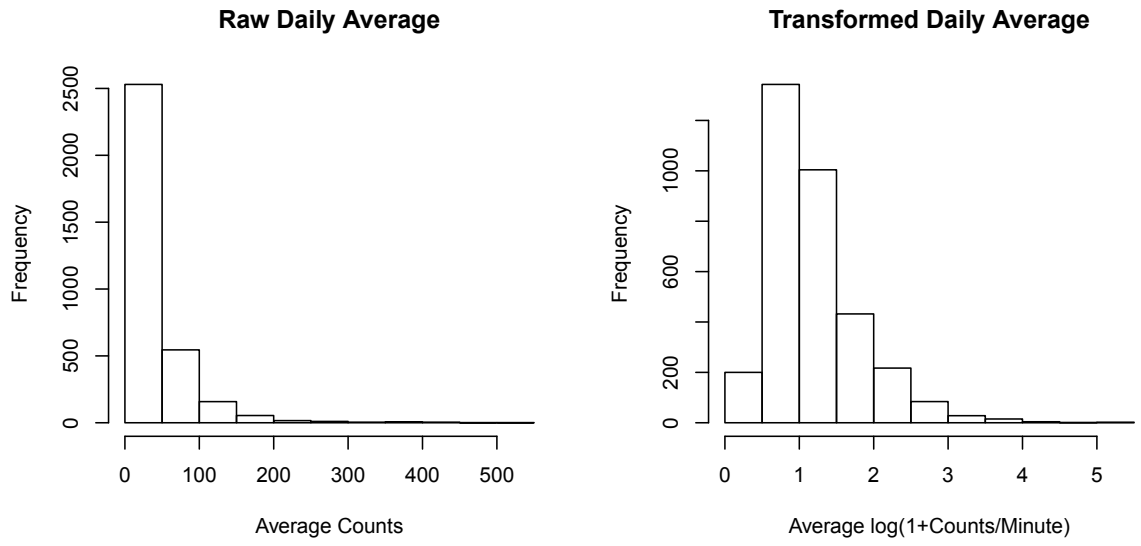
high degree of skewness of activity counts. In this data, the majority of activity counts are close to 0 (median activity count across all days and phases is 0, 3^{rd} quartile is 11), while some are quite high (maximum observed minute level activity count is 27,130). Correspondingly, the distribution of daily average counts per minute is heavily right skewed. It is therefore common practice in the analysis of count actigraphy data to use the transformation: log(1+counts/minute) rather than the raw data (Schrack et. al, 2013; Steeves et. al, 2014).

Applying the log transformation to the monkey activity data results in a much more symmetric distribution of daily average counts/minute. Figure 2.1 shows the change in distribution of daily average activity counts post log transformation. The histogram on the right (log transformed counts) exhibits meaningfully less skew than the histogram on the left (raw average daily counts per minute). As a result, we can more reasonably apply statistical tests that rely on the normality assumption to compare the transformed activity counts. To be precise, let y_{ijk} be minute level activity for monkey i, day j and minute k. Additionally, define $\widetilde{y}_{ijk} = \log(1 + y_{ijk})$ We can define the following quantities:

$$\text{Average Daily Activity Counts}_{ijk} = \bar{y}_{ij} = \frac{1}{1440}\sum_{k=1}^{1440} y_{ijk}$$

$$\text{Average Daily log(1+Counts/Minute)}_{ijk} = \bar{\widetilde{y}}_{ij} = \frac{1}{1440}\sum_{k=1}^{1440} \log(1 + y_{ijk}) = \frac{1}{1440}\sum_{k=1}^{1440} \widetilde{y}_{ijk}$$

Figure 2.1: *Comparison of Transformed Daily Average Counts/Minute*



Another prominent feature of this data is missingness at the day level. Table 2.1 lists the number of days for which there is data for each monkey during each of the 5 phases of the experiment. The order of the monkeys is organized by NTF treatment group assignment. As mentioned previously, 5 monkeys were not included in the NTF portion of the experiment, and thus were not assigned a treatment group. Correspondingly, there are 0 days of observations for each of these 5 monkeys during the three infusion phases. However, there are 0 days of observation for other monkeys during various phases of the experiment. For example, monkey 5 has 0 days of data during Pre-MPTP, Post-MPTP and Infusion 3.

In analyzing the data, special attention was paid to missing days as informative missingness could be problematic. Most of the missing days of data resulted from: pen changes, equipment (actigraphy device) malfunction or forced sedation for operations relating to the experiment. That is, within a treatment phase, the missingness does not depend on a monkey's behavior/activity/response to treatment. Therefore, it may be reasonable to assume that the missingness is missing at random.

Table 2.1: *Number of Days of Data for Each Phase of Experiment (by NTF Group)*

Treatment Group	ID	PreMPTP	PostMPTP	Infusion1	Infusion2	Infusion3
150 CDNF	5	0	0	24	1	0
	13	14	23	29	25	26
	14	14	25	29	26	14
	17	14	24	29	25	25
	19	15	25	25	26	37
	22	15	26	26	25	36
450 CDNF	1	14	19	24	20	23
	3	6	24	33	24	22
	18	14	24	30	16	25
	24	15	19	21	29	30
450 GDNF	2	15	18	26	22	24
	6	3	24	12	27	22
	12	15	25	29	21	25
	23	15	18	25	25	35
	35	14	26	0	0	31
N2	20	15	25	16	26	33
	26	15	26	25	25	36
	27	14	21	22	26	30
	28	0	0	0	0	25
	30	4	23	23	25	31
	31	14	24	34	24	28
N4	15	15	24	21	26	36
	25	14	26	34	26	30
	32	15	20	30	29	30
	33	14	23	27	24	32
	34	14	26	27	25	0
Vehicle	8	14	15	29	26	25
	11	14	26	29	27	27
	16	15	18	20	26	36
	29	14	23	0	14	5
Not Assigned	4	14	19	0	0	0
	7	4	14	0	0	0
	9	14	28	0	0	0
	10	14	28	0	0	0
	21	14	28	0	0	0

2.2 Visualizing Activity

The first step to analyzing complex activity data is to decide how best to visualize the data. Considering the structure of the experiment and heterogeneity of missing days, the most natural way to view the activity data is by considering each of the five phases of the experiment individually. Here we took three main approaches that will be detailed in the remainder of this section.

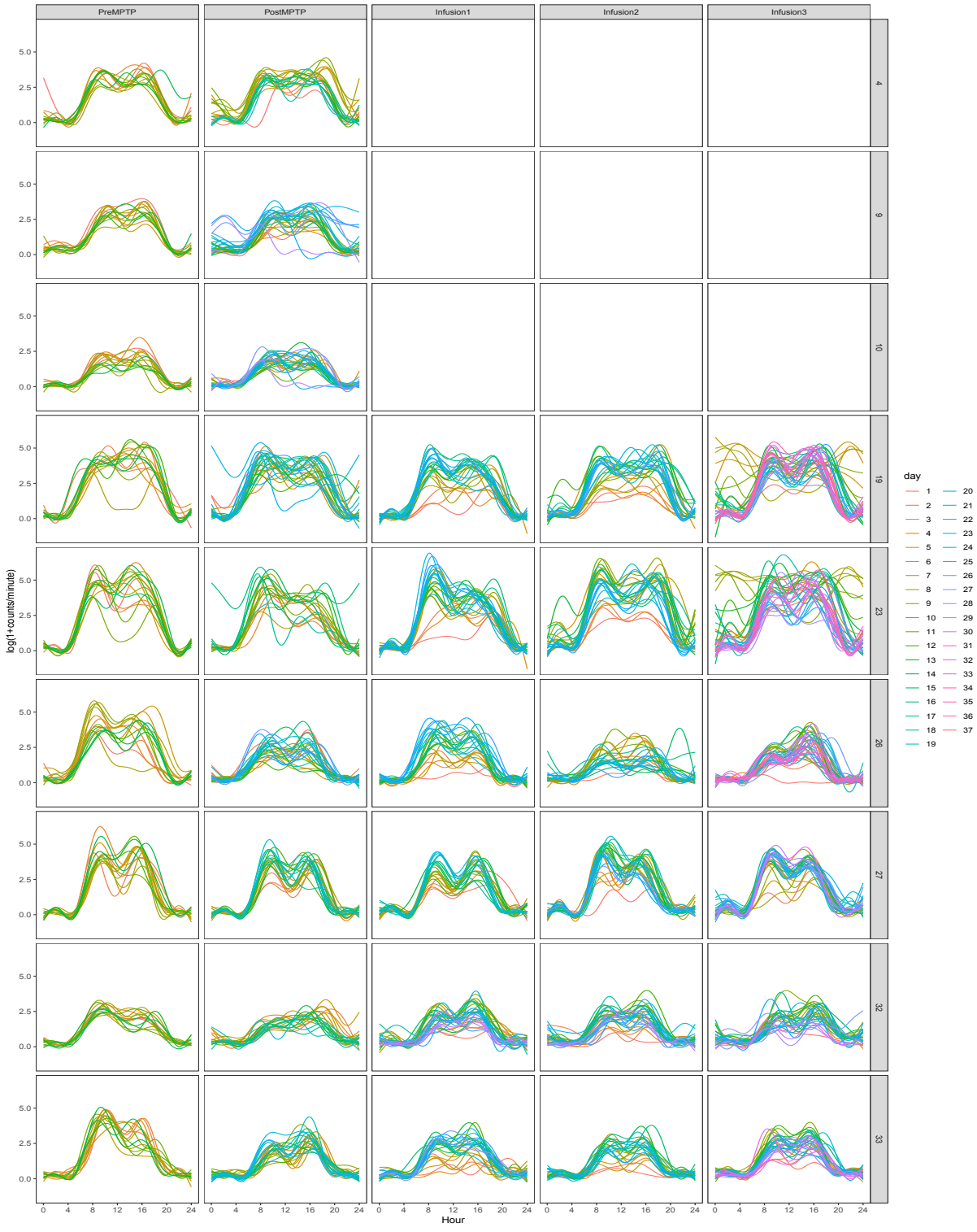
2.2.1 Individual Daily Curves

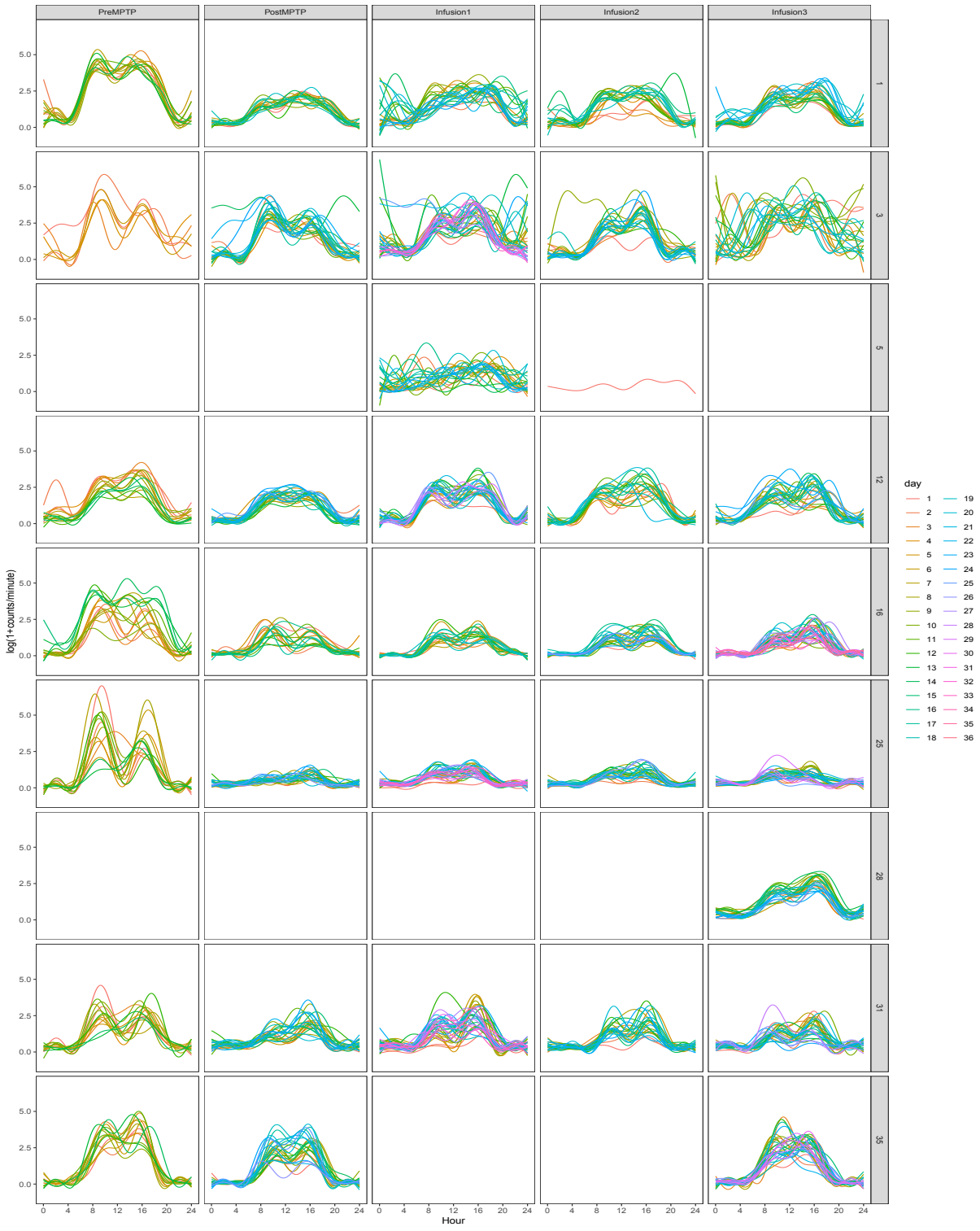
Activity can be thought of as a smooth process throughout the day, but at the minute level resolution, the data for an individual day can be quite noisy. To smooth the noisy data, generalized additive models were fit for each monkey-day using the mgcv package in R by regressing each day's log transformed activity counts on a smooth function of the integers 1:1440 (Wood, 2014). Generalized additive models will be described in more detail later in this document. Figure 2.2 displays the smoothed daily activity curves for all monkeys, in all five phases of the experiment. Each row corresponds to a single monkey while each column represents a phase of the experiment. Within each panel, different colors indicate different days. Plots with no lines indicate no data is available for those monkeys during the associated phase (column) of the experiment.

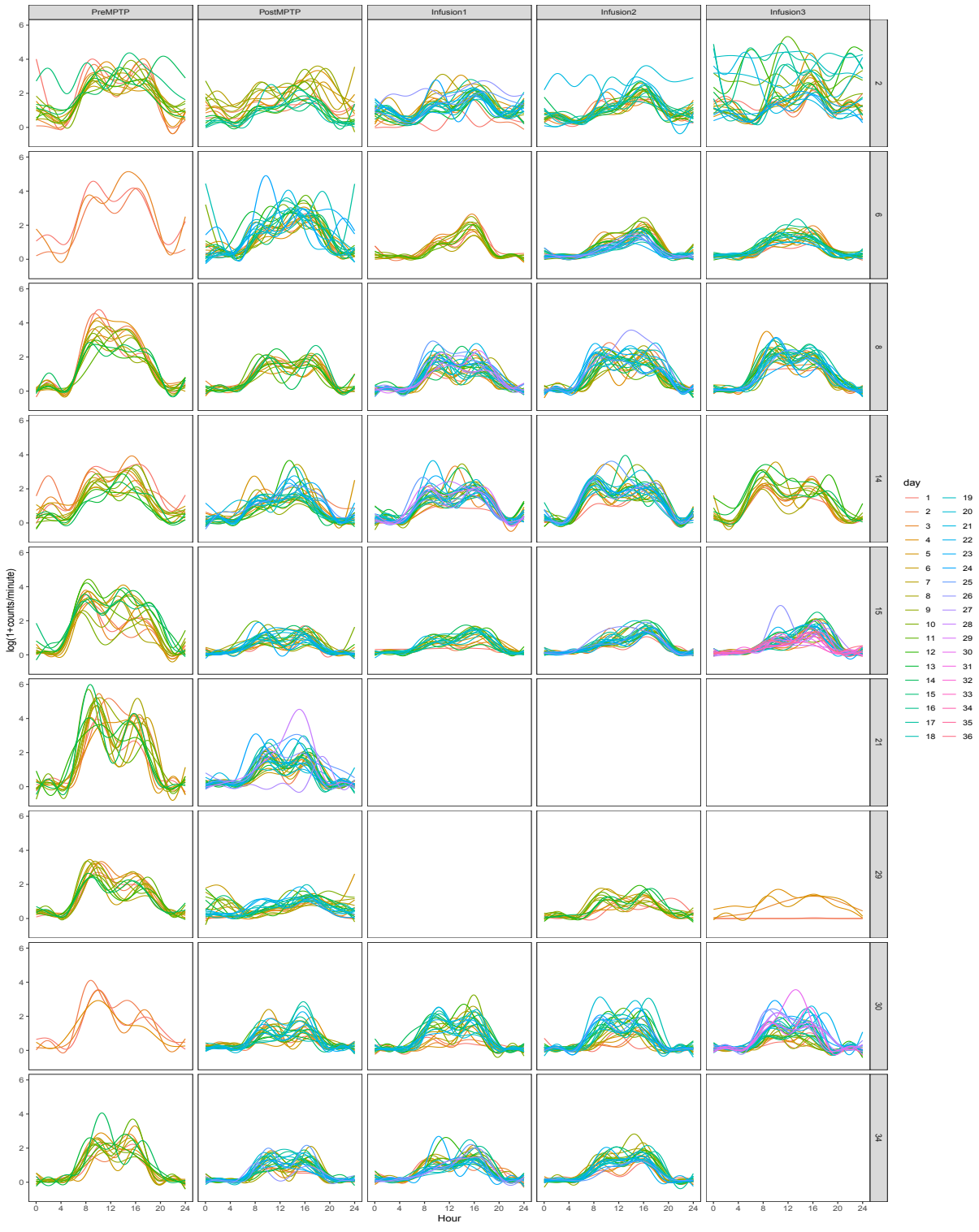
Examination of the first column (baseline) makes clear that in some monkeys, there is considerable day-to-day variability at baseline. In other monkeys, the smoothed activity pattern appears relatively consistent across days. Another visible feature observed in the daily curves of most monkeys at baseline is the bi-modal peaks located close to 10am (hour 10) and 4pm (hour 16). Furthermore, monkeys generally seem to have relatively stable sleep patterns (indicated by activity near 0 activity during night times) at baseline with the exception of monkeys 2, 3, 13, 14, 16.

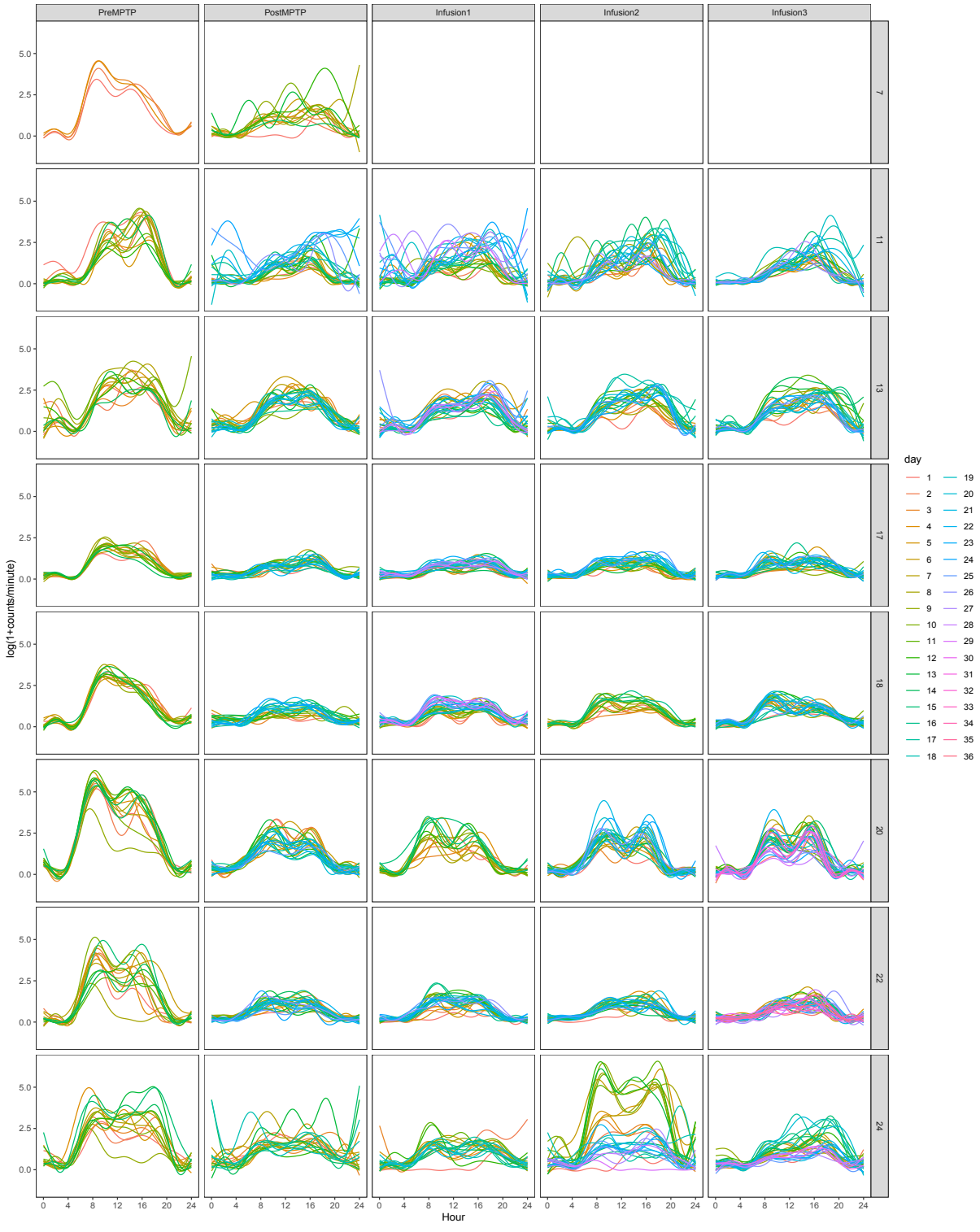
Looking at the MPTP phase (column 2), there is a decrease in daily activity during the activity periods for the day in most monkeys. In some monkeys this decrease is quite dramatic (1, 15, 25, 29). Additionally there appears to be increased nighttime activity in a number of monkeys. This may indicate a disruption of the natural sleep cycle due to MPTP administration. There may also be increased variability within/across days and a potential disruption of the circadian (bi-modal peak) rhythm observed at baseline. However, due to the clustering of lines in these plots, it is difficult to confidently assert such a difference exists from visual examination alone.

Figure 2.2: *Smoothed Individual Curves*




PreMPTP
PostMPTP
Infusion1
Infusion2
Infusion3
1
3
5
12
16
25
28
31
35
log(1+counts/minute)
Hour
day


PreMPTP
PostMPTP
Infusion1
Infusion2
Infusion3
2
6
8
14
15
21
29
30
34
log(1+counts/minute)
Hour
day
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36


PreMPTP
PostMPTP
Infusion1
Infusion2
Infusion3
7
11
13
17
18
20
22
24
log(1+counts/minute)
Hour
day

2.2.2 Individual Average Daily Activity

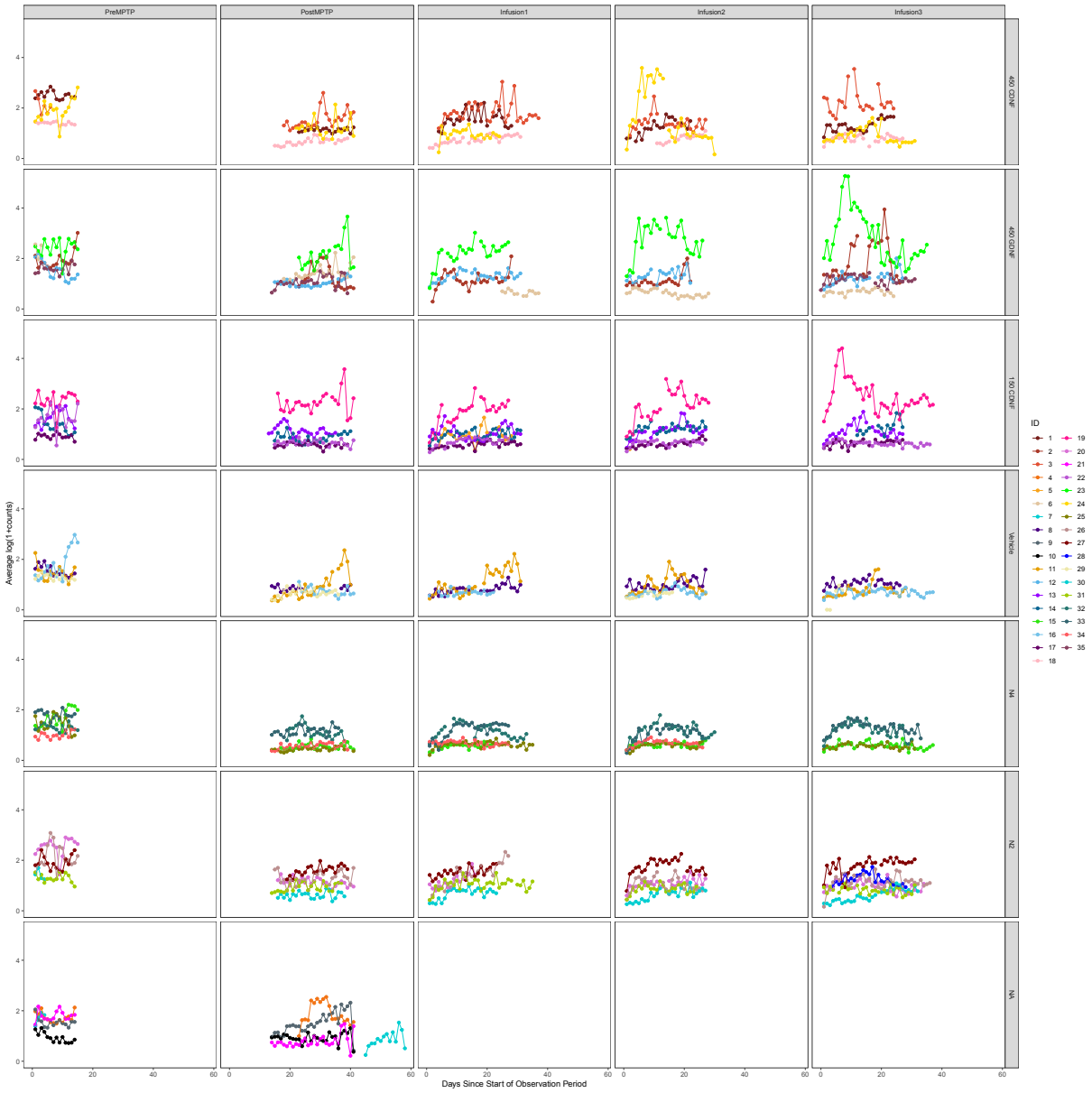
Although plotting the individual curves provides useful insights into the within/between monkey variability of activity curves, the sheer volume of visual data makes pattern recognition challenging. One strategy for reducing the visual complexity is to plot some daily level summary measure for each monkey in each phase of the experiment. To that end, Figure 2.3 examines the mean of the log(1+counts/minute), $\bar{\tilde{y}}_{ij}$, for each of the monkeys in each phase of the experiment. The layout for this figure is similar to that of Figure 2.2. Here, as with Figure 2.2, columns represent the five phases of the experiment. Rows, however, divide the monkeys into NTF treatment groups with the last row being those monkeys who did not receive a treatment assignment as they were not included in those phases of the experiment. Each line provides the trend for individual monkeys. Gaps in the lines indicate missing days.

The x-axis indicates the days since the start of the experiment phase. In the case of 'PreMPTP' (baseline) all monkeys start at 0 since this is their observation with no treatment. As mentioned in Section 1.2, PostMPTP administration, monkeys were given approximately two weeks to recover before they were fitted with their accelerometer again. One monkey that was not assigned a treatment group was either not fitted with an accelerometer until ∼40 days after the surgery or the data quality prior to that day was inadequate for further analysis. In the NTF Infusion phase, however, monkeys were generally fitted with their accelerometers shortly following surgery. As such, it is unsurprising to see in a majority of monkeys a slight upward trend in activity following the start of each of these phases followed by a leveling out as monkeys recover from the anesthesia/surgical procedure.

From this figure using just the average daily log(1+counts/minute), it can be seen that, at the daily level, average overall activity within a day is relatively stable for most monkeys at baseline. In addition, most have $\bar{\tilde{y}}$ between 1 and 2.5 across the days in this

phase. While that is a relatively wide range, the distribution across treatment groups looks fairly similar. In the Post-MPTP phase (column 2) it is clear that most monkeys exhibit at least some decrease in activity. It is not immediately obvious from this figure how average daily activity changes during the NTF Infusion phases (columns 3-5) of the experiment with the exception of a few monkeys.

Figure 2.3: *Average Daily log(1+counts/minute) Over All Observed Periods*



2.2.3 Smoothed Population Curves

Having looked at two different visualizations for the longitudinal/cross-sectional activity patterns at the individual level, we turn our attention to population level activity patterns. To do this, we employ a P-spline (penalized B-spline) bi-variate smoothing technique to smooth across monkeys and days found in the 'refund' package in R (Crainiceanu et. al, 2013; Xiao, 2013). Essentially, a penalized B-spline is a piece-wise polynomial function that smooths data, but penalizes 'roughness' in a way that makes the curve less jagged. By using this bi-variate smoother, we are able to borrow information across monkeys and days to create smooth population level activity curves throughout each phase of the experiment.

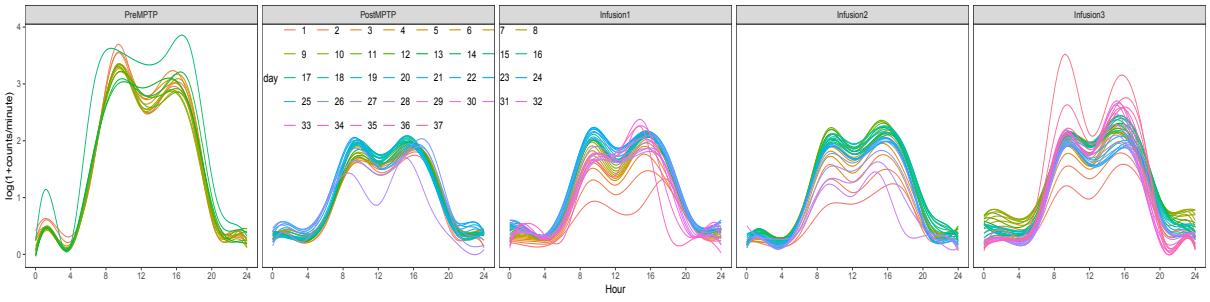
The result of this smoothing process is one smooth curve per day. The smoothing method was applied to each of the phases independently due to the unbalanced nature of the starting dates for each phase across monkeys. The result is five matrices of dimension $max(n_{ip}) \times 1440$ where i indicates monkey and p indicates phase of experiment. For instance, the monkey with the most days of observation in the Pre-MPTP phase is 15 days, so the resulting matrix will be 15×1440. Two approaches taken to plotting these curves are described below.

First, Figure 2.4, expresses the curves in a format similar to that employed in Figure 2.2. Each panel represents one phase of the experiment, with the process reading chronologically from left to right. Each curve is associated with a particular day indicated by color. Here the difference in overall activity level between Pre-MPTP and Post-MPTP phases is even more visually apparent than in the previous figures. It is also clear that at the population level there is relative consistency across days within each of these two phases with two potential exceptions. Furthermore, the bi-modal peak of daily activity with modes at approximately 10am (Hour 10) and 4pm (Hour 16) becomes clear. The magnitude of the peak in the morning also tends to be larger than that in the afternoon

Pre-MPTP.

From the three right most panels in Figure 2.4 corresponding to the NTF infusion phases, the recovery from anesthesia/surgery can be seen. The orange curves, corresponding the the earliest days in each phase, start low and increase until they seem to level out. It also looks like there are a few curves in phase 2 associated with the later days of that phase (color purple) where low average activity is observed. If, however, we look at the Infusion 2 column of Figure 2.2, we see that for those later days only a few monkeys are contributing data. Moreover, these monkeys have average activity lower than the other monkeys. A similar phenomena can explain the one unusually high daily activity curve in the Pre-MPTP phase (color green). Thus, we must be careful when interpreting these smoothed curves for days where there are relatively few monkeys contributing data.

Figure 2.4: *Smoothed Daily Activity Curves Over All Monkeys*



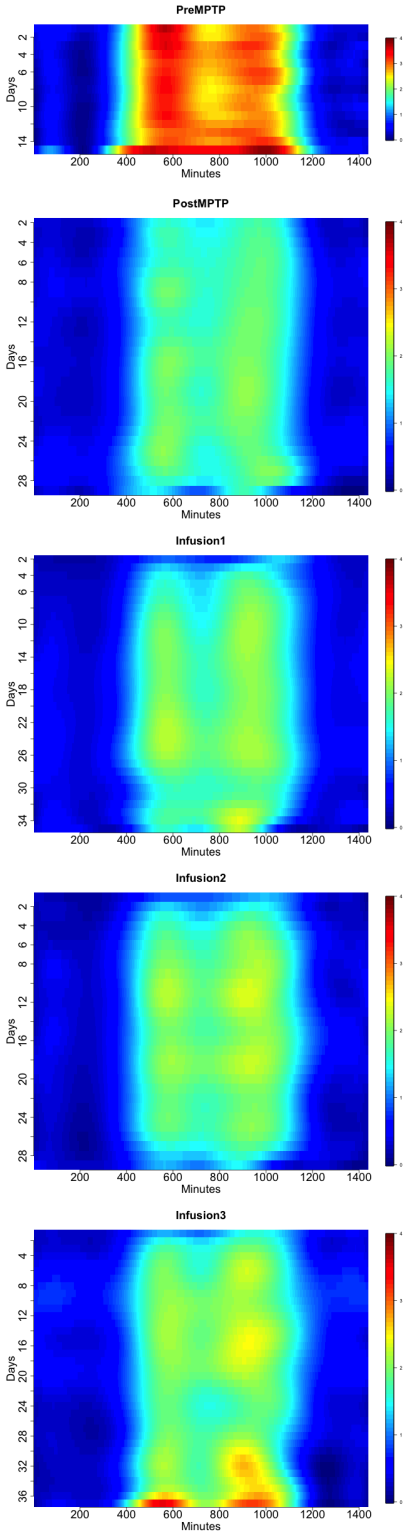
Even though Figure 2.4 allows for better visualization of patterns at the population level, longitudinal trends within phases are difficult to spot. To help with visualizing the longitudinal changes, heatmaps of the activity were created for each of the five phases as displayed in Figure 2.5. The Figure is organized into five heatmaps, corresponding to precisely one of the experiment phases. In each of the plots, the x-axis indicates time of day $1-1440$ minutes from left to right. The y-axis reads days from the start of the phase. Note that the origin (days=0) starts at the top of each plot rather than at the traditional

bottom. This allows for visual examination of the entire experiment longitudinally by reading from top to bottom down each of the five heatmaps sequentially.

Again, the drop in activity during the active period of the day comparing Pre-MPTP and Post-MPTP appears quite large. In addition, monkeys appear to have shorter activity periods as indicated by the narrower band of non-dark blue colors during the 6am-8pm (Minutes 360-1200) range. Interestingly, during the NTF infusion phases it appears that at the population level, the afternoon peak is larger than that of the morning. Suggesting that either the NTF treatment or residual effects from the MPTP may be affecting not just overall activity, but also activity patterns.

While there does appear to be some increase in overall activity from Post-MPTP period through the NTF infusion periods (indicated by more yellow during active periods), it does not appear to be dramatic with the exception of days 32-38 in the Infusion 3 phase. Referring back to 2.3 it becomes clear that only 2-4 monkeys are providing information for these days and two of those monkeys have average activity far above the rest of the monkeys.

Figure 2.5: *Heat Map of Actigraphy Data - All Phases*



Perform bi-variate smoothing in R

1. Set up the data as presented below (note the data can be set up in a number of ways, but to make the method clear columns will be referred to by name). Here we will only show the first 6 columns of the dataframe, but the remaining columns 7 : 1443 will contain the activity counts for minutes 4 : 1440. Each row contains an entire day of activity data for one monkey (MnkyID), which phase of the experiment the data is associated with (Infusion) and the associated days since start of that phase (day). Assume this object is a dataframe called 'data'.

day	Infusion	MnkyID	Minute1	Minute2	Minute3
1	PreMPTP	1	0	0	0
2	PreMPTP	1	0	0	0
3	PreMPTP	1	22	11	0
4	PreMPTP	1	0	0	0
5	PreMPTP	1	0	0	0

2. Execute the following code:

```
require(refund); require(fields)
## Subset the data for the phase you want to view
tmp <- subset(data,data$Infusion=='PreMPTP')
## get the log(1+counts/minute) in a new matrix
X   <- as.matrix(log(1+tmp[,paste('Minute',1:1440,sep='')]))
## n1, n2 are dimensions which to smooth over
n1  <- max(tmp$day);  n2  <- 1440
## x, z create the 'covariate' vectors
## indicating where on the grid to estimate
x   <- tmp$day/n1 - 1/2/n1;  z   <- (1:n2)/n2 - 1/2/n2
## apply the smoother
```

```
est <- fbps(X,subj=unlist(tmp$MnkyID),covariates=list(x=x,z=z),knots=8)

## extract the unique daily smoothed curves

hat <- est$Yhat[match(1:n1,tmp$day),]

## plot the heatmap (optional)

image.plot(1:n1,1:n2,hat,xlab='Days',ylab='Minutes')
```

3. Response to MPTP

3.1 Change in Mean Activity

The first and most natural question to ask regarding the effect of MPTP on monkey's activity relates to overall average daily activity $\left(\bar{\widetilde{y}}\right)$. Figures 2.3, 2.4 and 2.5 suggest that there is a decrease in overall activity on the magnitude of approximately $1-1.25$ on the log scale during peak activity hours. To quantify this difference, one could, within a monkey, take average across all days during Pre-MPTP and Post-MPTP phases separately and perform a paired t-test. A paired t-test is not necessarily the most powerful test as there is some loss of information when averaging across days within a monkey. One way to utilize all days of data for all monkeys while respecting the correlation structure between days within monkeys is to employ Generalized Estimating Equations (GEE). This method will be described in detail as we will use GEE to create a comprehensive model of monkeys' longitudinal activity throughout all phases of the experiment (Laird & Ware, 1982; Liang & Zeger, 1986).

In addition to population level changes in average activity, it is also of interest to see which monkeys had significant changes in their daily activity. To answer this question we use unpaired t-tests for each monkey individually comparing their average daily activity ($\bar{\widetilde{y}}$) Pre-MPTP to Post-MPTP. A Bonferonni correction is applied to the p-values resulting from these tests to correct for multiple testing (Dunn, 1961).

3.1.1 Generalized Estimating Equations

When working with repeated measures data (i.e. multiple observations for a single subject) one usually needs to account for the within subject correlation structure. Typically this problem occurs in the context of regression. In the case where the outcome (conditioned on individuals) is roughly Gaussian, a solution for the estimating marginal models proposed by Laird & Waire is to use a two-stage random effects model. Marginal models refers to the fact that the estimate of the outcome does not depend on the covariate history for previous time points. That is, marginal models answer the questions relating to population regression parameters.

The idea is to first model the regression parameters within each individual, and then model the variation between individuals. We will describe the process using a single random effect (random intercept). Let α_0 be the random intercept, describing monkey's different 'mean' activity at baseline. Next, let $\tilde{\bar{y}}_{ij}$ be the $(n_i \times 1)$ vector of daily average of log(1+counts/minute) for individual i (where n_i is the number of repeated observations of monkey i). Further, let X_i be the $(n_i \times p)$ design matrix that links the population level effects β to y_{ij}. The two stage iterative estimating process is described below:

(1) Consider the model: $\tilde{\bar{y}}_{ij} = \alpha_i + \beta X_j + \epsilon_{ij}$ where $\epsilon_{ij} \sim N(0, R)$ (normally distributed with mean $\mathbf{0}$ and covariance matrix R). Given the most recent estimates of α_0 and D, estimate the variance/covariance matrix R and solve for β.

(2) Assume $\alpha_i \sim N(0, D)$ where D is of dimension 1×1 (or, more generally, number of random effects $\times$ number of random effects). If there are more than one random effect, they are usually assumed to be independent. Given the estimate of β from step (1), update estimates of α_0 and D.

Laird & Waire (1986) provide a set of estimating equations used to iteratively update this process. Although the covariance matrix R may be left fully unspecified, this results

in a loss of one degree of freedom for each additional pairwise correlation estimated. Therefore, it is generally used only when the number of repeated measures is small relative to the number of individuals. Even in situations where it is reasonable to leave the within subject correlation fully unspecified, the algorithms employed in statistical software packages may not converge to a solution.

In this case, there are more repeated measures than monkeys and is therefore not feasible to use an unspecified covariance structure. In addition, we believe it is reasonable to assume an exchangeable correlation for both Pre-MPTP and Post-MPTP phases as we don't observe any obvious changes in ordering of average activity for monkeys. Furthermore, assuming a correctly specified mean model, GEE is robust to misspecifications of within-subject correlation structure and the estimates for β are consistent (Fitzmaurice et. al, 2011).

Liang & Zeger provided a set of estimating equation procedures (generalized estimating equations) that extended this framework to the generalized linear model (1986). These estimating equations have been implemented in R using the package geepack (Højsgaard et al, 2006; Yan & Fine, 2004; Yan, 2002). We will use this package in the remainder of the paper whenever we fit a model using GEE.

3.1.2 Average Daily Activity

Utilizing generalized estimating equations and assuming an exchangeable correlation structure between days within monkeys (Equation 3.1), a statistically significant decrease in average daily log(1+counts/minute) was found (95% CI: [-0.825,-0.408]). To test for changes in average activity level temporally across the day, generalized estimating equations were again used to examine changes at the hourly level (i.e. 24 tests, one test for every hour of the day). Adjusting for multiple testing using the Bonferonni correction, a significant decrease in activity was found for all hours including and between 6am and

7pm. Table 3.1 contains the adjusted 95% confidence intervals for these hours with red indicating the difference for that hour is statistically significant.

$$\tilde{\bar{y}}_{ij} = \alpha_i + \beta X_j + e_{ij} \tag{3.1}$$

Where X_j is a $n_i \times 1$ vector with $x_j = 0$ if day j is Pre-MPTP and $x_j = 1$ if day j is Post-MPTP for monkey i:

$$\alpha_i \sim N(0, D), \quad e_{ij} \sim N(0, R), \quad R = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & \ddots & \rho & \rho \\ \vdots & & & \ddots & \vdots \\ \rho & \dots & \dots & \rho & 1 \end{bmatrix}$$

Table 3.1: *Change in Hourly Population Average Activity: This table contains the estimate and 95% Bonferroni corrected confidence intervals for the change in activity (Post-MPTP administration - Pre-MPTP administration) averaged at each of the 24 hours of the day. Estimates/hours colored red indicate statistical significance.*

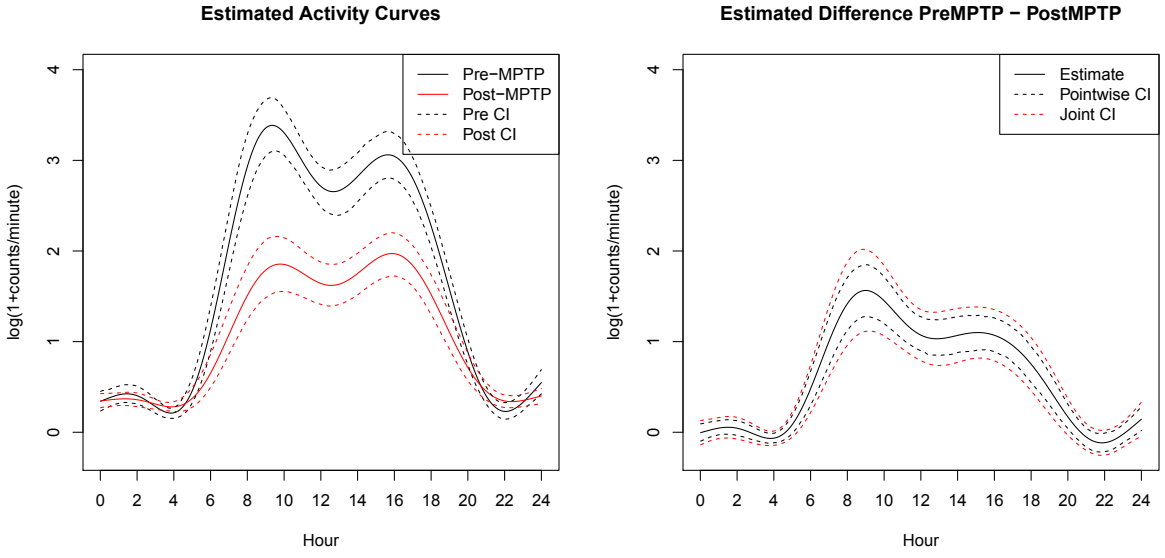
Hour 1	Hour 2	Hour 3	Hour 4	Hour 5
-0.01 (-0.14, 0.12)	0.00 (-0.12, 0.12)	-0.05 (-0.20, 0.09)	0.03 (-0.05, 0.11)	-0.02 (-0.10, 0.06)
Hour 6	Hour 7	Hour 8	Hour 9	Hour 10
0.03 (-0.14, 0.21)	-0.87 (-1.55, -0.20)	-1.43 (-1.97, -0.89)	-1.36 (-1.87, -0.86)	-1.27 (-1.72, -0.81)
Hour 11	Hour 12	Hour 13	Hour 14	Hour 15
-1.40 (-1.84, -0.95)	-1.30 (-1.68, -0.91)	-0.95 (-1.28, -0.61)	-0.99 (-1.34, -0.64)	-1.13 (-1.43, -0.82)
Hour 16	Hour 17	Hour 18	Hour 19	Hour 20
-1.05 (-1.39, -0.72)	-0.99 (-1.31, -0.67)	-0.94 (-1.33, -0.56)	-0.69 (-1.01, -0.37)	-0.23 (-0.50, 0.04)
Hour 21	Hour 22	Hour 23	Hour 24	
0.03 (-0.16, 0.21)	0.03 (-0.14, 0.19)	0.04 (-0.07, 0.15)	0.01 (-0.13, 0.15)	

To get a continuous estimate of the difference Pre-MPTP to Post-MPTP, a bootstrap

procedure was employed. By bootstrapping individual monkeys, we were able to obtain confidence intervals for the average activity curves Pre-MPTP to Post-MPTP as well as a confidence interval for the difference. The left panel of Figure 3.1 shows the confidence intervals for the Pre-MPTP and Post-MPTP activity curves while the right plot indicates the confidence interval for the difference (both point-wise and joint). The joint confidence interval was obtained by multiplying the point-wise standard errors by the .95 quantile of the maximum of the standardized absolute difference of observed difference and mean difference for all time points $t \in (1, \ldots, 1440)$.

Figure 3.1 and Table 3.1 agree in significance of the difference in activity, but vary slightly on the width of the confidence intervals. From both the table and the figure, it is clear the largest difference occurs at approximately 9am, corresponding the the morning peak. Although monkey activity is slightly higher Post-MPTP during most of the sleeping hours, this difference is not statistically significant.

Figure 3.1: *Bootstrapped Confidence Intervals for Activity Curves*



3.1.3 Average Activity for Individual Monkeys

Applying unpaired t-tests for changes in the mean daily activity for individual monkeys, 9 of the 33 monkeys did not show a significant decrease in average daily log activity counts after adjusting for multiple testing using the Bonferroni correction (3, 4, 6, 9, 10, 19, 23, 30 and 32). Three of these 9 monkeys (4, 9 and 10) were excluded from the NTF phase of the study. Of the remaining 6 monkeys, 5 had significant decreases in activity using p-values when not adjusting for multiple testing. Table 3.2 provide the Bonferroni adjusted 95% confidence intervals for change in average daily log(1+counts/minute) for each of the 33 monkeys for which we have both Pre-MPTP and Post-MPTP data. Red confidence intervals indicate a statistically significant difference.

Table 3.2: *Subject Change in Daily Average Activity: This table contains the* 95% *Bonferroni corrected confidence intervals for the change in activity (Post-MPTP administration - Pre-MPTP administration) for each of the subjects in the study. Subjects are identified by their ID numbers. Estimates/IDs colored red indicate statistical significance.*

1	2	3	4	6	7	8
(-1.54,-1.20)	(-1.15,-0.12)	(-1.34, 0.20)	(-0.26, 0.55)	(-4.36, 2.50)	(-1.60,-0.12)	(-0.90,-0.45)
9	10	11	12	13	14	15
(-0.37, 0.29)	(-0.24, 0.22)	(-1.00,-0.10)	(-0.88,-0.11)	(-0.85,-0.16)	(-0.93,-0.12)	(-1.51,-0.77)
16	17	18	19	20	21	22
(-1.71,-0.45)	(-0.46,-0.22)	(-0.84,-0.63)	(-0.51, 0.36)	(-1.77,-1.01)	(-1.26,-0.74)	(-1.41,-0.53)
23	24	25	26	27	29	30
(-0.82, 0.32)	(-1.22,-0.18)	(-1.33,-0.67)	(-1.35,-0.31)	(-0.64, 0.00)	(-0.77,-0.47)	(-1.77, 0.16)
31	32	33	34	35		
(-0.56,-0.14)	(-0.40, 0.07)	(-0.85,-0.47)	(-0.59,-0.27)	(-0.75,-0.28)		

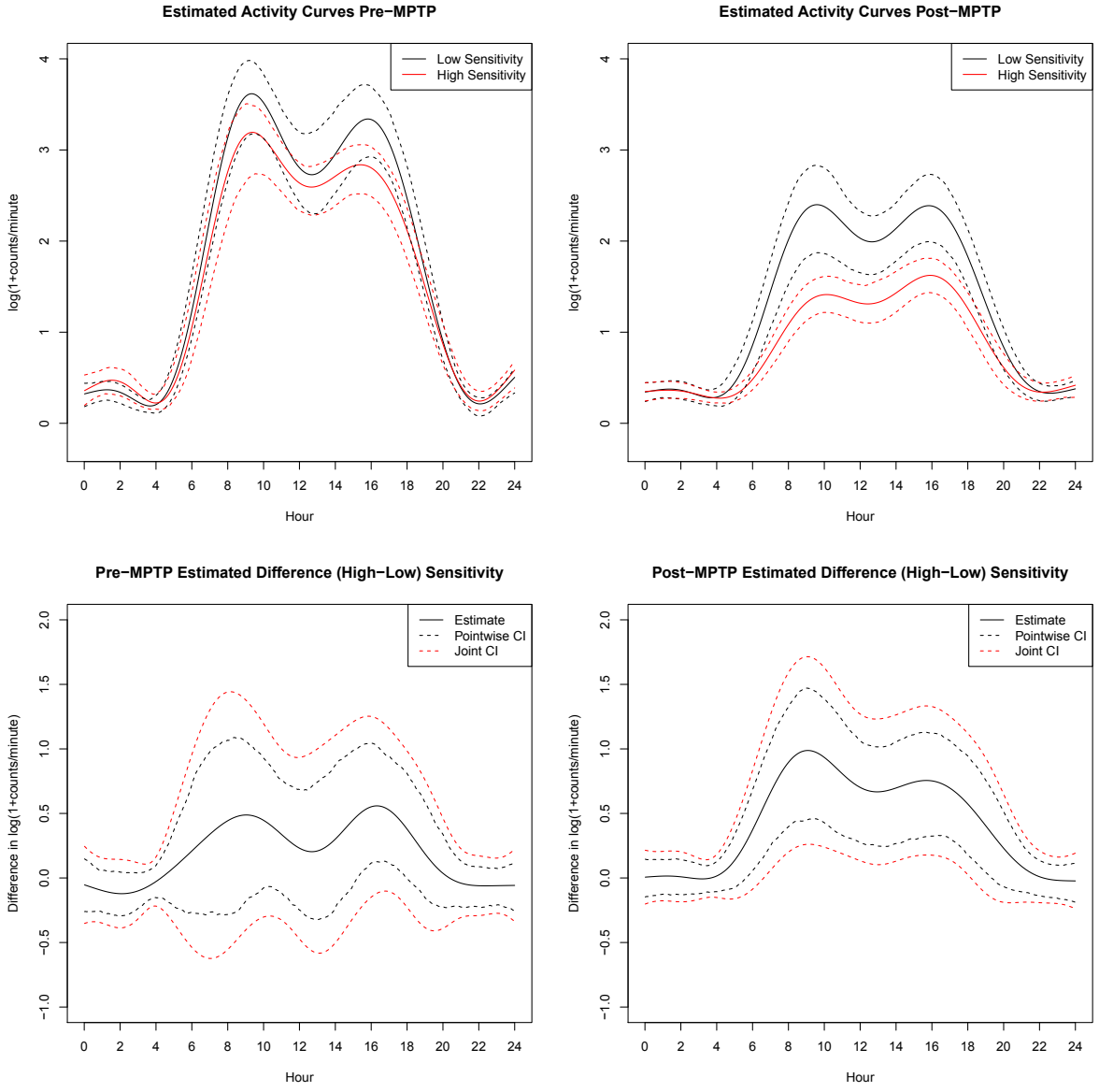
3.1.4 Association of Parkinson's Rating Scale and Activity

The outcome most recognized for clinical significance in the assessment of severity of Parkinson's symptoms is the Parkinson's rating scale. It is known that activity is as-

sociated with severity of Parkinson's symptoms. Moreover, it is known that in humans increased activity is associated with better long term health outcomes. However, it is unsettled as to whether activity predicts severity of Parkinson's symptoms or if activity is just diminished as a result of the onset of Parkinson's.

Figure 3.2 plots the activity curves of monkeys dichotomized by the median rating scale. Those in the upper 50^{th} percentile are those monkeys who responded more strongly to MPTP (negatively) and can be thought of as being highly sensitive to MPTP. The left column shows the average activity curves (upper left) and the estimated difference throughout the day (lower left) with confidence intervals obtained by bootstrapping monkeys Pre-MPTP. The right column shows the activity and difference Post-MPTP. Although highly sensitive monkeys have lower daytime activity, this difference is not significant Pre-MPTP. Post-MPTP administration, however, we do see a significant difference in average daily activity during the peak activity hours.

Figure 3.2: *Activity Profiles of Monkeys with High vs Low MPTP Sensitivity*

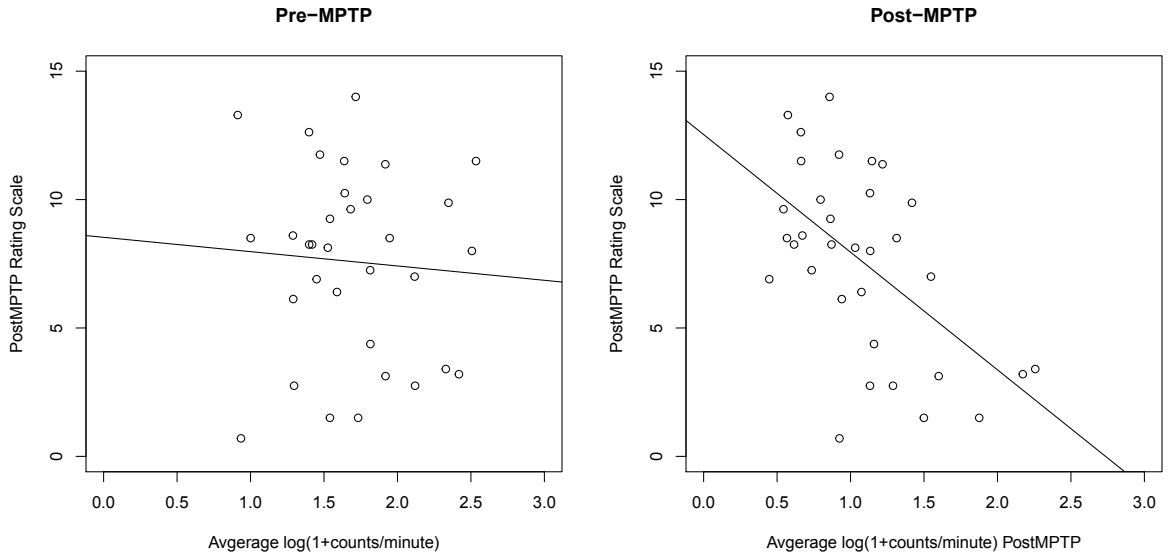


Although we dichotomized the monkeys by rating scale, the metric is thought of as continuous clinically. As a result, there is no well defined, clinically meaningful cut-off point for measuring severity. Respecting this structure, the association (correlation) of rating scale with both average daily log(1+counts/minute) pre-MPTP and post-MPTP was calculated. Figure 3.3 plots the average daily counts against the Parkinson's rating

scale (Pre-MPTP on the left, Post-MPTP on the right). This plot shows that average activity pre-MPTP (that is, before onset of Parkinson's symptoms) exhibits only a slight negative correlation with rating scale (-0.07, p-value=0.707). Once Parkinson's symptoms have developed, we do see a significant negative correlation (-0.57, p-value=0.001) with rating scale.

These results imply that average activity prior to developing Parkinson's symptoms (pre-MPTP administration) is not predictive of the severity of Parkinson's symptoms in response to MPTP. This lack of association between average overall activity and MPTP response motivates a more sophisticated analysis of activity patterns.

Figure 3.3: *Correlations between Rating Scale and Pre/Post MPTP Average Daily Activity*



3.2 Patterns of Activity

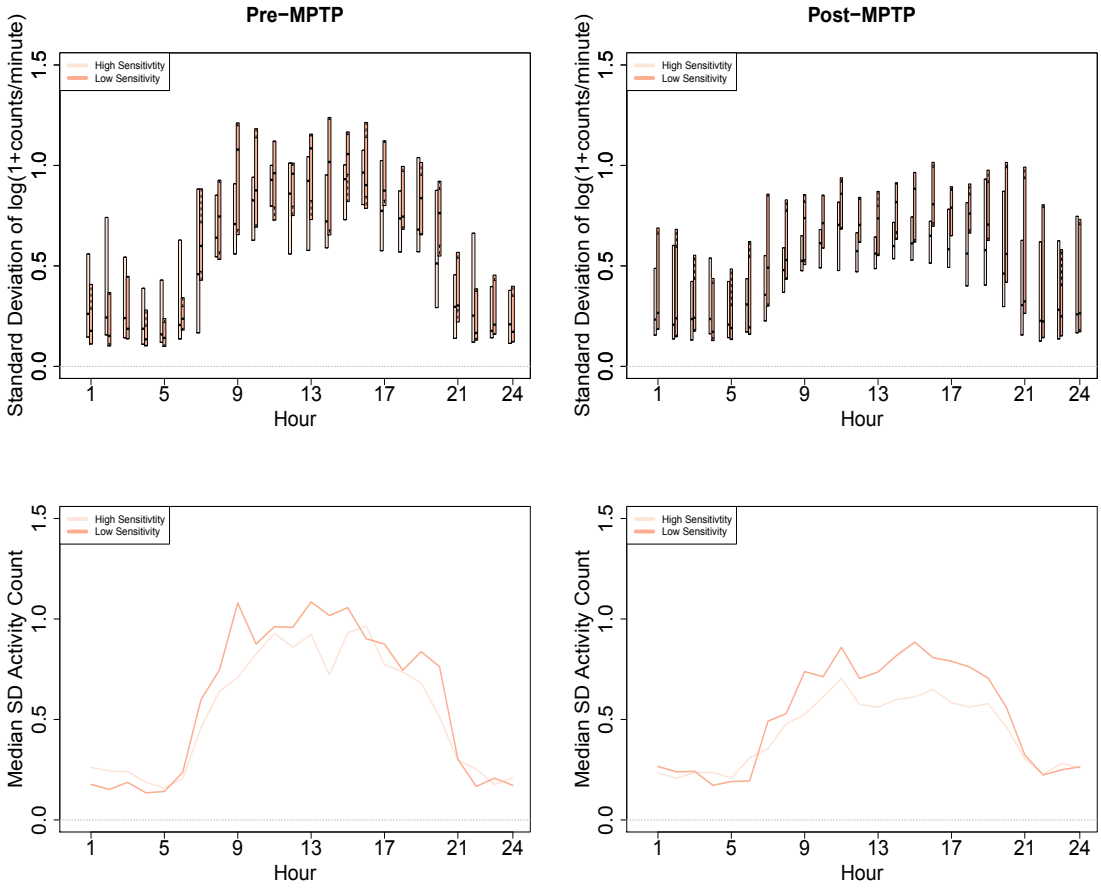
3.2.1 Day-to-Day Standard Deviation in Activity Counts

In analyzing patterns of activity, the investigation is to consider the within monkey day-to-day standard deviation in log transformed activity counts. Using the high/low

sensitivity dichotomy presented in the previous subsection, the day-to-day standard deviation of log transformed activity counts (averaged at the hourly level) was calculated for both pre-MPTP and post-MPTP phases of the experiment. Figure 3.4 displays the distributions for standard deviation of hourly average log activity counts for the two groups of monkeys. The first row of the plot provides the inner quartile range for the distribution of standard deviations for each hour of the day. The second row plots the median of the distribution of standard deviations for each hour of the day. The left column (both first and second row) refers to the pre-MPTP phase while the right column explores the post-MPTP phase.

Looking at the left column of Figure 3.4, there does appear to be some differentiation between the low and high sensitivity groups during the waking hours, with the largest differences occurring at hours 9, 14 and 18-20. None of these differences, are significant for $\alpha = 0.05$ using two sample t-tests. Again, since our dichotomizing of the monkeys is somewhat arbitrary, the correlation between rating scale and standard deviation (using pre-MPTP phase activity counts) was calculated. The periods with highest correlation are hours 15, 20 and 12 ($\hat{\rho} = -0.32, -0.22, -0.21$, respectively). These are also the periods in which the largest differentiation in Figure 3.4 was observed between high and low sensitivity monkeys. Although these correlations were found not to be significant using permutation test ($\alpha = 0.05$), the magnitudes suggest activity patterns may predict response to MPTP.

Figure 3.4: *Day-to-Day Standard Deviation of Log(1+Activity Counts/Minute)*



3.2.2 Functional Principal Component Analysis (fPCA)

Having established variability in activity as at least a potential predictor of activity, a reasonable place to look for more complex patterns is in the functional principal components of monkey activity. The reason for this is twofold: first, activity data is functional data (thus motivating the use of fPCA versus PCA). In general, principal component analysis allows one to examine high dimensional sources of variability (patterns) summarized in relatively few dimensions. Said differently, the method allows for meaningful reduction of dimensionality while maintaining focus on the sources of highest variability in the data. In particular, we are interested in seeing if the dominant circadian patterns

elicited from the first K principal components are predictive of response to MPTP in monkeys. Therefore fPCA was performed on the Pre-MPTP actigraphy data only.

Defining fPCA

Adopting the notation described by Ramsay & Silverman (2005), we describe how to perform traditional principal component analysis using minute level actigraphy data for multiple subjects over multiple days, Then we describe what differentiates classical PCA from functional PCA. Imagine there are p covariates (here $p = 1440$ for every minute in the day). Arrange the data in an $(N \times p)$ matrix where N is the total number of monkey days Pre-MPTP. Each row of this matrix corresponds to one day of data for one monkey and each column represents one minute of data ordered chronologically.

The key idea is to find the linear combination ($\xi_{\mathbf{1}}$) of the p predictors that explains the maximum amount of variance in the data. The goal is then to find another linear combination that explains the maximum amount of remaining variance as possible, subject to the constraint that this next linear combination ($\xi_{\mathbf{2}}$) is orthogonal to the first. This process is repeated to generate new linear combinations that explain successively less variance in the data. These ξ are referred to as principal components. Generally just the first few principal components are used as they usually explain a sufficient percentage of variation in the data. Moreover, interpretation of higher order principal components becomes increasingly difficult as their effects are dominated by the preceding principal components.

The covariance matrix of the demeaned data $\mathbf{V} = \frac{1}{n}\mathbf{X}'\mathbf{X}$ can be decomposed as $\frac{1}{n}\mathbf{V}\boldsymbol{\Sigma}^{\mathbf{2}}\mathbf{V}'$. As such, in practice PCA is performed by first calculating the singular value decomposition of the demeaned data matrix $\mathbf{X} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}'$ due to computational concerns. Columns of $min(p, N-1) \times p$ matrix $\mathbf{V}$ correspond to a 'principal component'.

The main distinction between traditional PCA and fPCA is the smoothing of the

covariance matrix. The rationale for smoothing in fPCA is the belief that true functional data is smooth. To perform this smoothing, one must first choose a basis function. Since we can consider this type of activity to be periodic, a natural choice of basis uses Fourier series of sufficient complexity. Suppose we have a series of $\mathcal{V}$ Fourier series denoted $\phi(\mathcal{V})$. Let $x(s)$ be the function measuring daily activity. Assuming x is periodic (daily) this function can be expanded as a Fourier series with coefficients $c_{\mathcal{V}} = \int x\phi_{\mathcal{V}}$: $x(s) = \sum_{\mathcal{V}} c_{\mathcal{V}}\phi_{\mathcal{V}}(s)$. Ramsay & Silverman describe the following steps to perform Functional PCA (2005):

(1) The coefficients $\mathbf{c}_i$ are calculated for every day of monkey activity.

(2) Smooth the covariance matrix of Fourier coefficients $\mathbf{c}_i$.

(3) Perform standard PCA on the smoothed covariance matrix.

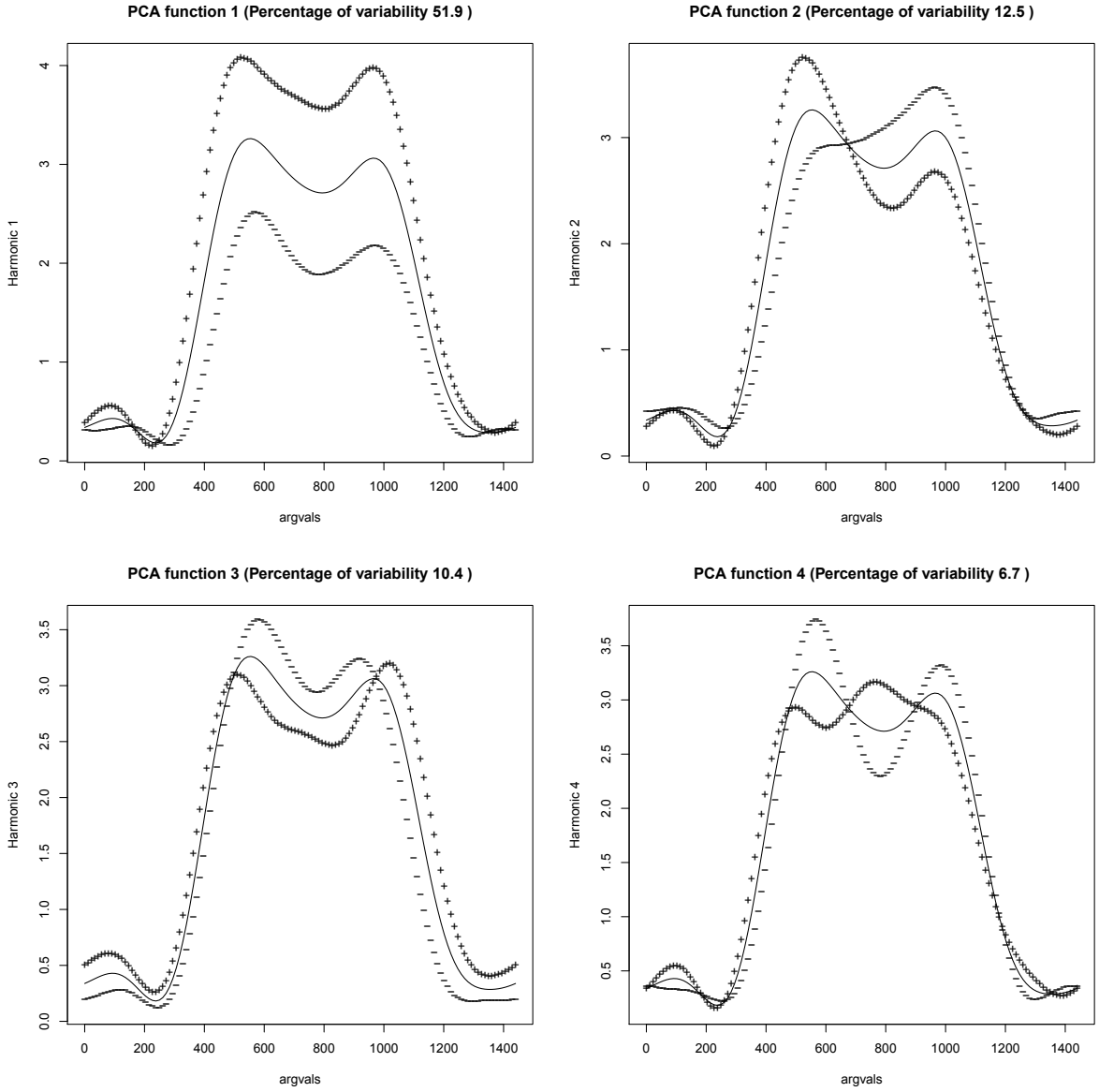
(4) Apply the same smoothing operation as in step (2) to the eigenvectors resulting from step (3) and normalize so these vectors are orthonormal.

(5) Compute the principal component function ξ by taking the inner product of the normalized, smoothed eigenvectors from step (4) $\mathbf{y}$ and $\phi(s)$.

Results of fPCA

Applying functional principal component analysis to the daily activity pre-MPTP (using the fda package in R), Figure 4.4 was produced. Here, the x-axis is denominated in minutes instead of hours. The y-axis measures log(1+counts/minute). These plot shows the first four harmonics which together explain 81.5% of the variability in the data. The dashed (+) and (−) lines indicate the effect of a $+/-2$ standard deviations from the mean principal score loading. Each harmonic can be thought of as defining successively important features of the daily activity curve in monkeys.

Figure 3.5: *First 4 Harmonics Resulting from fPCA Applied to Pre-MPTP Activity*

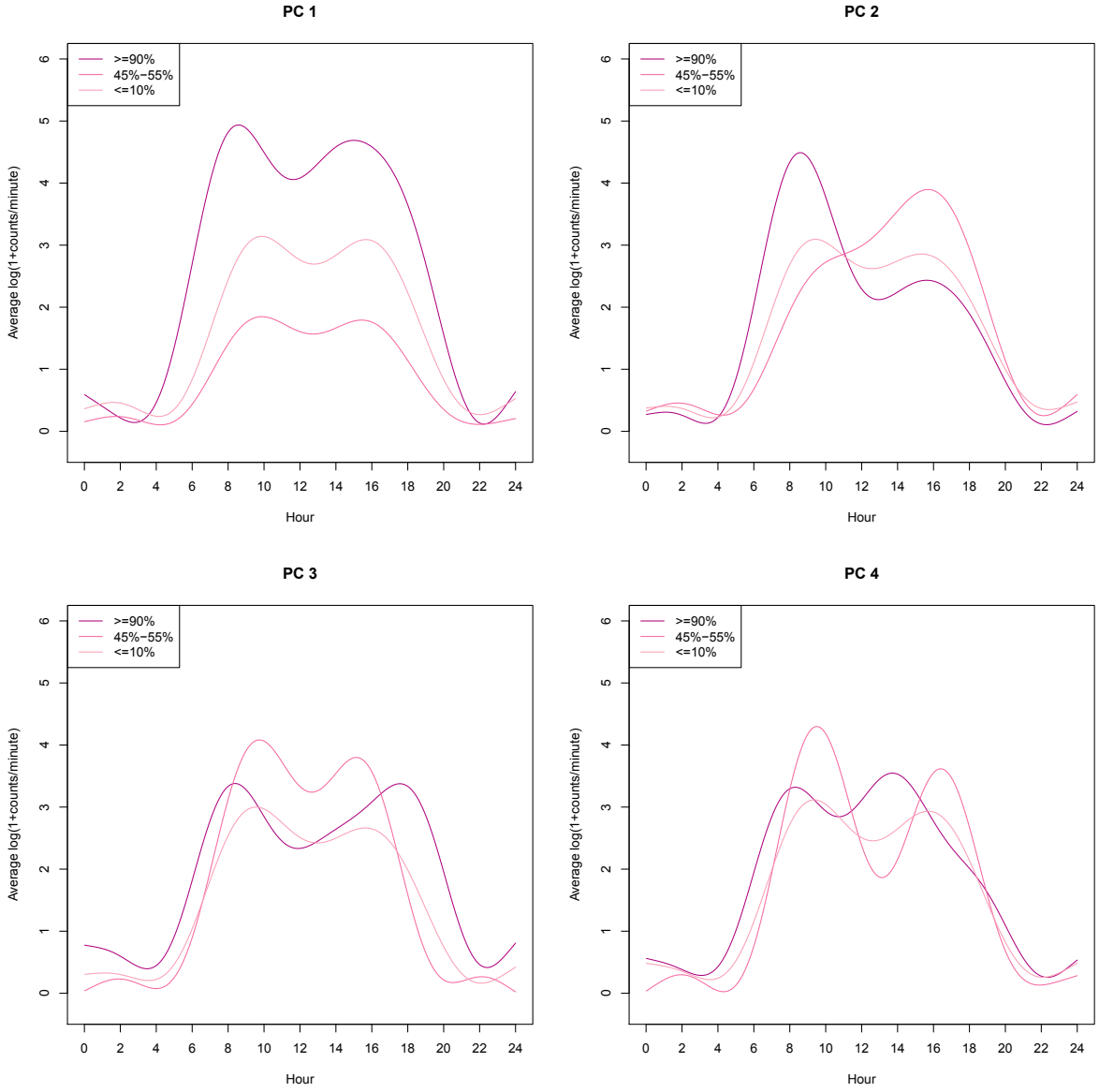


The weightings of PC scores (eigenscores) for the first harmonic can be interpreted as comparing overall activity during the active hours of the day. This mean shift in the activity curve accounts for most of the variation in the data (51.9%). Specifically, comparing a day weighted highly on the first principal component would have higher average level of activity than a day less highly weighted. The second harmonic compares activity

in the morning relative to activity in the afternoon. The third harmonic compares activity during sleeping hours versus activity during midday. The fourth harmonic compares monkeys with high activity during the two peak activity hours with monkeys who have higher activity during the afternoon dip (napping time).

In addition to plotting the components, Figure 3.6 plots the smoothed average activity profiles of monkey days corresponding to the 0.0-0.1, 0.45-0.55 and 0.9-1.0 quantiles of the distributions of the eigenscores for the first four principal components. By and large the activity profiles mimic those seen in Figure 3.5, however the activity profiles corresponding to the tail quantiles for PC3 and PC4 differ slightly from their patterns from the plot of the harmonics. This is likely due to the influence of PC1 and PC2.

Figure 3.6: *Activity Profiles by Quantile of PC Score*



Correlations were calculated with both mean and standard deviation of daily principal component scores. The results can be seen in Table 3.3 (a). The correlations are not particularly large nor are they statistically significant. To understand more fully the associations, each of the means and standard deviations for the first four principal component scores were plotted. Figure 3.7 shows the standard deviation of the third and

fourth principal component scores plotted against post-MPTP rating scale. A moderately strong negative association exists between the standard deviation of both principal component scores and rating scale, if monkeys with rating scale below 2 are excluded. In fact, these monkeys (4, 9, and 10) were excluded from the NTF infusion portion of the study based on their lack of response to MPTP. When these monkeys are excluded, the correlations of rating scale with the standard deviation of principal components 2, 3 and 4 increase considerably and become statistically significant for $\alpha = 0.05$ (Table 3.3 (b)).

Table 3.3: *Correlations between Principal Component Scores (Mean & SD)*

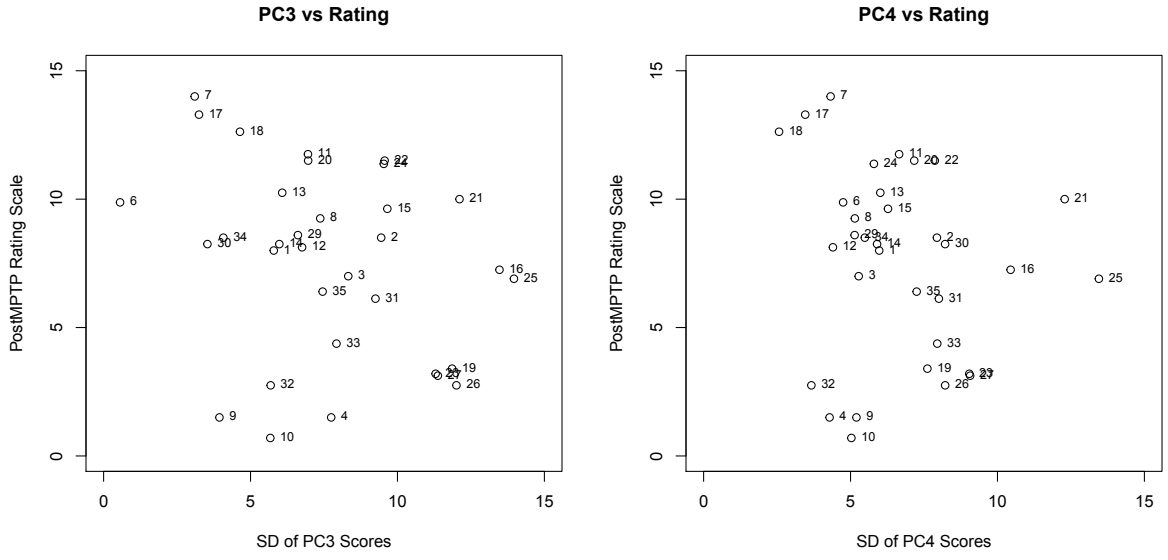
(a) Full Data

	PC1	PC2	PC3	PC4
Mean Score	-0.09	0.14	-0.00	0.21
SD of Score	-0.04	-0.17	-0.29	-0.16

(b) Monkeys with Rating > 2

	PC1	PC2	PC3	PC4
Mean Score	-0.27	0.00	0.00	0.23
SD of Score	-0.20	-0.38	-0.47	-0.36

Figure 3.7: *Association between Rating Scale and SD of PC Scores*



Performing fPCA in R

Using functions from the fda package, fPCA can be performed quite easily in R. The steps described above are shown below.

```
## get just the activity counts and take the log
X             <- data[,paste('Minute',1:1440,sep='')]; X <- log(1+X)

## Create the fourier basis with correct period and 11 basis functions
basis         <- create.fourier.basis(c(0, 1440), nbasis=11, period=1440)

## Creates smoothing penalty (harmonic)
harmonic_Lfd  <- vec2Lfd(c(0,(2*pi/1440)^2,0), c(0, 1440))

## Aggregates basis, smoothing penalty and smoothing parameter
harmonic_fdPar<- fdPar(basis, harmonic_Lfd, lambda=1e5)

## Smooth the covariance matrix and extract the 'fd' object
smoothfd      <- smooth.basis((1:1440)-.5, t(as.matrix(X)),basis)$fd

## Perform the fPCA on the smoothed covariance matrix and
## extract the first K=4 harmonics, then plot them
fPCA          <- pca.fd(smoothfd, nharm=4, harmonic_fdPar)
plot(fPCA)
```

3.2.3 Scalar on Function Regression

So far the methods used to predict response to MPTP shown have been increasingly complex, taking into account more complicated features of the activity profiles of monkeys. To model response to MPTP in a fully functional fashion, we turn back to the idea of generalized additive models (GAMs) first mentioned in Section 2.2.1. In that section, generalized additive models with a single smooth term were used to smooth activity curves. To show how this is possible, we first need to define generalized additive models. Using this structure, we can non-parametrically model linear predictors that are a function of some covariate (x) and time (t).

Defining (f)GAMs

Recall: the generalized linear model assumes that an outcome y is distributed with some mean μ and variance, such that for some exponential family link function η such that $\eta(\mu_i) = \sum_{j=1}^{J} \beta_j x_i$. The generalized additive model does not place the parametric restriction on the association between each predictor (β) and the outcome (Hastie & Tibshirani, 1986, 1990). To be precise, we can define the generalized additive model as follows:

$$\eta(\mu_i) = \sum_{i=j}^{J} f_j(x_{ij}) \tag{3.2}$$

Where f_j is any smooth function of the J covariates. Note that these functions f_j can be parametric in the generalized additive model. The smoothing of activity profiles performed in Section 2.2.1 resulted from models defining each day of activity as a 1440×1 response vector regressed on the vector $t = 1, 2, \ldots, 1440$. That is, for every monkey day we fit the model: $\log(1 + \text{counts}_t) = f_1(t)$ using the mgcv package. The fitted values from these models create a smooth activity curve.

In the mgcv package, the smoothing is done using penalized regression splines. The idea is to place a 'penalty' on jaggedness of a curve which can be expressed in terms of a smoothing parameter λ (Wood, 2006). Consider the simple case where the link function is the identity function ($\eta(\mu) = \mu$) and there is one smooth predictor s and one predictor x. The solution to linear regression is often expressed as the set of $\boldsymbol{\beta}$ coefficients that satisfies the minimization problem: $||\boldsymbol{y} - \boldsymbol{X\beta}||^2$. However, by placing no parametric assumptions on the association between x and y, the solution to the least squares problem would be to fit a curve that passes through every observed point. In most instances this would be over-fitting the model and yield little useful inference. One solution is to fit a smooth curve that places a penalty on the 'roughness' of the curve. To add the penalty, the minimization problem can be reformulated as:

$$||\boldsymbol{y} - \boldsymbol{X\beta}||^2 + \lambda \int_0^1 \left[f''(x)\right]^2 dx$$

$$\text{or, equivalently } ||\boldsymbol{y} - \boldsymbol{X\beta}||^2 + \lambda \boldsymbol{\beta^T S \beta}$$

Where $\boldsymbol{S}$ is a matrix of known spline coefficients.

The solution to the penalized least squares problem is then $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X^T X} + \lambda \boldsymbol{S}\right)^{-1} \boldsymbol{X^T y}$. In practice, this is fit using Penalized Iteratively Re-weighted Least Squares (P-IRLS) as a result of the complexities that arise when estimating a GAM where the link function is not the identity function. By framing the problem in this fashion, the model fit is not particularly sensitive to choice of knots so long as the number is sufficiently large (Ruppert, 2002). Rather, the smoothness of the curve is mostly affected by choice of λ. The gam function in the mgcv package uses cross validation to choose λ.

Functional Generalized Additive Models (fGAM)

In order to move from to the functional generalized additive model case, we now assume that we again have some outcome Y_i for subject i and some functional predictor $X_i(t), t \in 1:1440$. In the situation where the link function is the identity function, McLean et. al (2012) proposed the following model:

$$\mathbb{E}(Y_i|X_i) = \theta_0 + \int_{t\in 1:1440} F(X_i(t), t)dt \tag{3.3}$$

In effect, this model allows for the outcome to depend on both the level of the predictor X and the time at which this value is observed. That is, this model allows for the possibility that high activity at 1:00am has a fundamentally different association with the outcome than low activity at 1:00am which may in turn have a different effect than low activity at 11:00pm.

The estimation method proposed by McLean et. al utilizes the mgcv package in R and employs tensor product smooths of penalized B-splines (2012). Given some spline bases for activity (B^X) and time (B^T) where B^X and B^T have J, K knots respectively, Equation 3.3 can be expressed using the spline expansion:

$$\theta_0 + \int_T F\{X_i(t), t\}dt = \theta_0 + \sum_{j\in J}\sum_{k\in K}\theta_{j,k}\int_T B^X\{X_i(t)\}B^T(t)dt \tag{3.4}$$

Where $\int_T B^X\{X_i(t)\}B^T(t)dt$ can be estimated using numerical integration procedures. In analyzing the data, the effect of different bases on our results was examined. The three bases explored are as follows: tensor product smooths of P-splines, tensor product smooths of cubic regression splines, and thin plate splines.

To fit these models, two approaches were used. In the first, the log(1+counts/minute) for each minute $(1, \ldots, 1440)$ was averaged across days during the PreMPTP phase for

each monkey individually. This results in a vector of average log activity counts for each minute $1, \ldots, 1440$ (X_i) for each monkey. The resulting X matrix is of dimension 33×1440. Then, the Parkinson rating scores (Y_i) were regressed on these average activity profile using the gam function. This requires the construction of two additional matrices indicating to the gam function the time indices as well as the quadrature weights. The method and its justification is described in the online supplementary materials to McLean et. al (2012).

The second approach utilized the full range of data available for monkeys. To do this, the same response vector Y was used (33×1). However, instead of a regression on activity averaged across days, each monkey's activity was placed in 'wide' format. This regression requires construction of time and weight matrices that differ in important ways from the first approach. Following the presentation of the results is a description of how to fit these models in R. This approach makes the assumption that days are exchangeable and, assuming k days for each subject, Equation 3.3 now becomes:

$$\mathbb{E}(Y_i|X_{i,k}) = \theta_0 + \int_{t\in 1:1440} F(X_{i,k}(t), t)dt \tag{3.5}$$

Results of fGAM

Figures 3.8 and 3.9 show the results of the two approaches fitted models. Both figures present the estimated linear predictor, the estimated joint effect of activity and time, using three different choices of basis: tensor product smooths of P-Spline basis, tensor product smooths of cubic splines, smooths of thin plate splines.

Figure 3.8 plots the linear predictor for approach 1 (mean activity profiles) while Figure 3.9 plots the linear predictor for approach 2 (using all days). These plots can be interpreted as follows: consider a single day of activity curve for one monkey. If that curve were plotted on the (x, y) coordinates of these heatmaps, at each point there would be

an associated value of the linear predictor $Z_{x,t}$ (indicated by color intensity). The mean of these values amounts to the 'effect' of the linear predictor for that monkey curve. The addition of this term and the model predicted intercept is equal to the predicted response.

From both figures, we see that employing tensor product smooths of either P-splines or cubic splines results in a similar linear predictor. These plots indicate that lower activity early in the day and higher activity later in the day are associated with lower Parkinson's rating scales.

These results are in line with the prior results showing negative correlation with variability in average activity for hours 12, 15 and 20. That is, monkeys that occasionally either didn't nap midday or stayed active until later in the day would have lower predicted rating scores using these models. Additionally, activity at or near the average activity levels during the late morning/midday provide little to no information about the rating scale, in line with the previous findings regarding average activity.

Figure 3.8: *fGAM Estimated Linear Predictor Using Mean Activity Profiles*

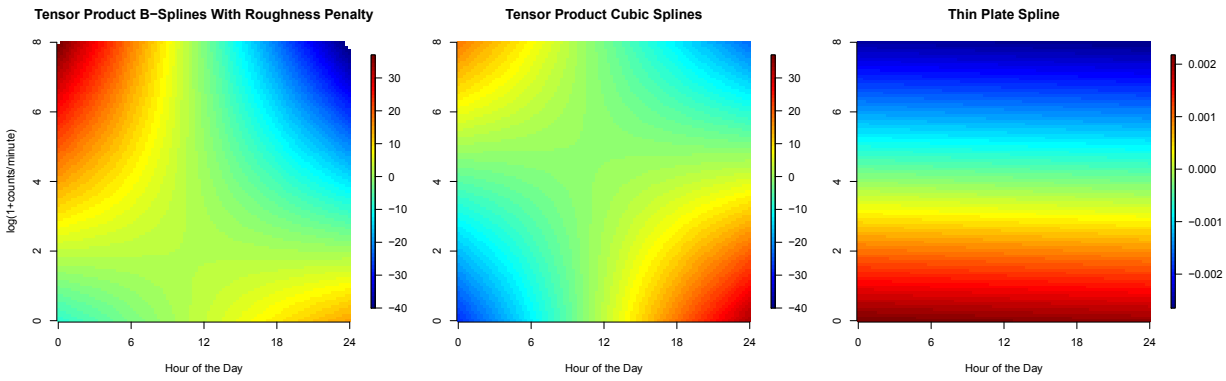
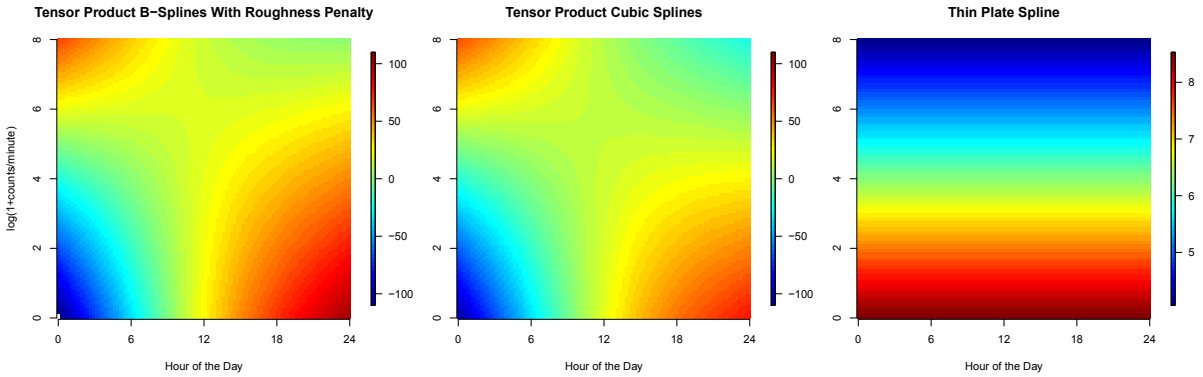


Figure 3.9: *fGAM Estimated Linear Predictor Using All Days*



Using thin plate regression splines to estimate the surface yields a plane that is essentially flat in the time direction. Interestingly, when we estimate rating scale using all days of activity, the predictions reduce to exactly the predictions made by the linear regression of rating scale on daily average activity (Table 3.4).

More than likely this is a result of scaling, since thin plate splines can are sensitive to scale while the tensor product smooths utilized in the gam function are scale invariant. For our estimation, we re-scaled time to be between 0 and 1. As a result, the effect of activity completely dominates when using thin plate regression splines. Note that using properly scaled covariates, thin plate regression splines have been shown to have some desirable properties. Although this thesis does not explore the effect scaling might have on the results, that could be a future application.

Lastly, It's important to remember when evaluating these plots that not all points on the grid are well populated. For instance, there are very few observations of activity greater than four log(1+counts/minute) before 3am. Estimates outside this range are then necessarily extrapolations beyond the observed data. Also, note that within each figure, the color intensity scales are the same except for the thin plate spline plot.

The question then becomes: what do we gain using these increasingly complex models of activity to predict Parkinson's outcomes? To answer that question, the sum of squared

residuals (SSR) for multiple models was calculated. The SSRs are displayed in Table 3.4. From this table, we see that there is an improvement in MSE over linear regression (1) in every model except fGAM using thin plate splines (6) for the reasons stated previously. Comparing fGAM using the mean activity profiles (2/3), to the fGAM models using all days of data (5/6), we see a sizable improvement in MSE.

Ultimately, a linear regression of Parkinson's rating scale on the standard deviation of the third principal component scores (8) actually performs the best in terms of SSR. However, the SSR for this model and models (5) and (6) are roughly comparable. Importantly, the difference in predictive power for fGAM models using all days versus using the average daily profile further supports the notion that there is more signal present in the variation between days than in average activity. Despite this improvement in SSR, none of the models presented here meet the threshold of statistical significance for $\alpha = 0.05$.

Table 3.4: *Sum of Squared Residuals for All Models*

Model	Sum of Squared Residuals
(1) Linear Regression on Average Daily Activity	428.47
(2) fGAM Mean Profiles (P-Splines)	422.73
(3) fGAM Mean Profiles (Cubic Splines)	422.73
(4) fGAM Mean Profiles (Thin Plate)	425.64
(5) fGAM All Days (P-Splines)	398.91
(6) fGAM All Days (Cubic Splines)	400.54
(7) fGAM All Days (Thin Plate)	428.47
(8) Linear Regression on SD of PC3	394.39

Simulation Study

As stated previously, it is unclear whether our inability to detect a statistically significant association between 'patterns' of activity and Parkinson's' rating score is due to a lack of power (small sample size) or a true lack of association. To more thoroughly explore the nature of this problem, a simulation study was performed where the true association

between activity and outcome was precisely the function estimated using tensor product smooths of P-splines (Figure 3.9, left panel).

Essentially, daily activity curves were generated from a 'population average' activity curve with noise. The population curve was created using a mixture of normal distributions and chosen to closely mimic the average population activity curves during the pre-MPTP phase of this experiment. Six different scenarios were examined using different number of subjects/monkeys (n = 33, 100), different numbers of days of observation (J = 1, 10) and different amounts of 'noise' added to to the simulated Parkinson's' scores ($\sigma_1^2 = 2.5, \sigma_2^2 = 1$). For every scenario we examined both in sample Mean Squared Error (MSE) and out of sample mean squared error predicted on 33 new subjects (MSE $= \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y_i)^2$). For the simulations with multiple days of observation, we also calculated the MSE for predictions using fGAM on the average activity profiles.

Figure 3.10: *Randomly Generated Activity Data vs Actual Data*

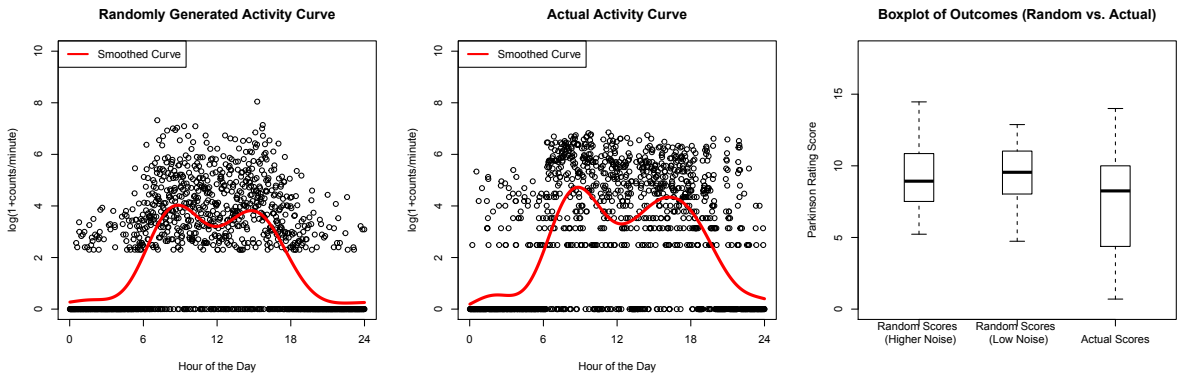


Table 3.5: *Quantiles of Activity for One Simulation of 33 Monkeys vs Observed Quantiles*

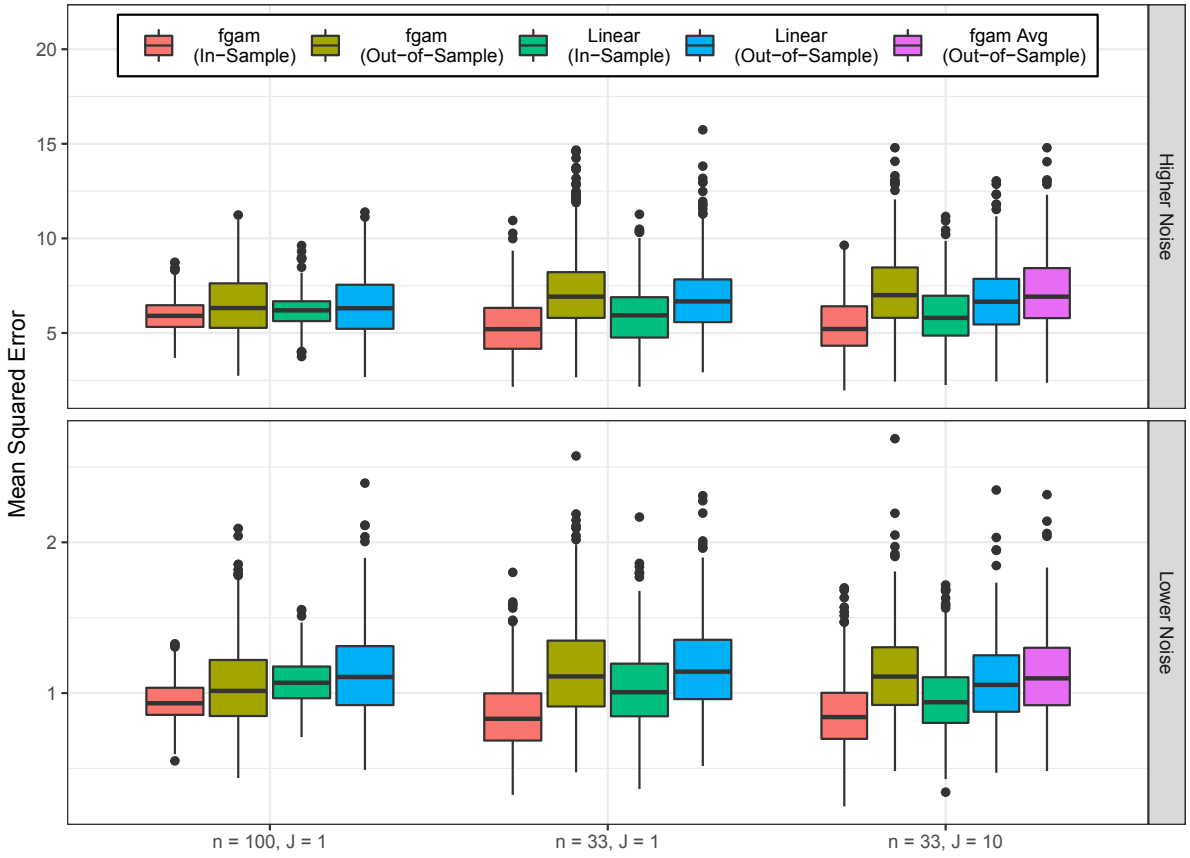
	Min	25th	Median	75th	90th	Max
Observed Activity Data	0.00	0.00	0.00	3.66	5.51	9.99
Randomly Generated Data	0.00	0.00	0.00	3.74	4.96	10.44

In assessing how closely the randomly generated data matches the observed data, a

number of plots and summary statistics were assessed. The left two panels of Figure 3.10 compares a randomly generated day of activity versus an observed day of activity. The smoothed curves appear similar in shape and the distribution of activity counts throughout the day also appear reasonably similar.

Table 3.5 Compares one distribution of activity counts over all days and monkeys for iteration of the simulation (n = 33) versus the distribution of all observed activity pre-MPTP. Again, the distributions look quite similar. The right panel of Figure 3.10 compares the distribution of observed outcomes (Parkinson's' rating scores) and our randomly generated outcomes. While the average of the randomly generated outcomes is higher, the distributions appear fairly similar for the scenario 'Higher Noise'.

Figure 3.11: *Distribution of Simulated Mean Squared Error*



Finally, Figure 3.11 shows the results of the simulation study. The top panel, corresponding to the higher noise model, shows that while in-sample performance is noticeably better using fGAM compared to linear regression (red vs. green), the out of sample prediction using fGAM is no better (gold vs blue), even though fGAM is the true model. This is true even with 33 subjects 10 days of observation. Importantly, these simulations do not propose monkey specific curves and the day-to-day variability within subjects is considerably lower than in the actual data.

Interestingly, however, in the bottom panel it can be seen that with 100 fGAM performs better out-of-sample than linear regression on the average activity counts. As a result of decreased noise (increased signal-to-noise ratio), fGAM performs better than linear regression for the out-of-sample prediction.

These results suggest several possibilities. It may be that there is truly no signal in activity patterns that predicts response to MPTP well. Alternatively, the signal may be present, but 33 monkeys may not provide sufficient power to detect the signal consistently with statistical significance. Future simulations might consider exploring the effect of monkey specific activity curves and imposing larger day-to-day variability in the activity counts that may capture more of the complex features of real-world activity data.

Performing fGAM in R

As mentioned previously, to fit fGAM models in R, we make use of the mgcv package. Both the simple (no repeated days) and the more complex (repeated days) will be explained. To fit this class of models, the data needs to be structured as follows:

- Y: An $n \times 1$ matrix of responses
- X: An $n \times J$ matrix of functional predictor values. For our simple case (models 1/2), this matrix would be 33×1440 ($xmat.wide$: For models (5/6), this matrix

should be $n \times 1440 \times$ max(number of days)). If not all subjects have the same number of days of data, missing values should be indicated by 0.

- *Tmat*: An $n \times J$ (*tmat.wide*: $n \times$ max(number of days)) matrix of time indices. For convenience, we re-scale time to be on the $[0, 1]$ scale, but this is not a requirement.

- L: An $n \times J$ (*lmat.wide*: $n \times 1440 \times$ max(number of days)) matrix of weights. This matrix should have rows that sum to 1. So, in the simple case, every entry in the matrix would be $\frac{1}{1440}$. In the scenario where there are multiple days of observation, each row needs to be $\frac{1}{1440} \times$(number of days of observation for that subject). Crucially, entries corresponding to missing functional covariate values need to be 0.

```
library(mgcv)


## fit the single day per subect models
fit  <- gam(y~te(X,Tmat,by=L,bs='ps') ,method='REML') ## P-splines
fit2 <- gam(y~te(X,Tmat,by=L),method='REML')          ## Cubic regression splines


## fit multiple days models using the same two bases
fit.full  <- gam(y~te(xmat.wide,tmat.wide,by=lmat.wide,bs='ps'),method='REML')
fit.full2 <- gam(y~te(xmat.wide,tmat.wide,by=lmat.wide),method='REML')


## Plotting the linear predictors
xind <- seq(0,8,len=100)
tind <- seq(0,1,len=100)
est.full <- predict(fit.full, type="iterms",
                            newdata=data.frame(xmat.wide=rep(xind, length(tind)),
```

```
                                        tmat.wide=rep(tind, each=length(xind)),
                                        lmat.wide=1))
est.full <- matrix(est.full,ncol=length(xind),nrow=length(tind),byrow=TRUE)
image.plot(est_full,xlab='Time',ylab='log(1+counts/minute)')
```

4. NTF Infusion Effect

4.1 Posthumous Brain Cell Counts

Following the observation period for infusion 3, the monkeys were sacrificed and instigators performed autopsys. From their examination of the deceased monkeys' brains, they determined that only 4 treatment groups seemed to have a nueroprotective effect using some measure of brain cell count relating to doperminergic neuron function. These neurotrophic factors were: 150 CDNF, 450 GDNF, N2 and N4. As such, the analysis that follows will focus specifically on the monkeys in those 4 treatment arms versus the monkeys that received the vehicle treatment.

4.2 Change in Mean Activity Counts

To assess whether we see a significant treatment effect, we fit the following model 5.1 using generalized estimating equations. This model allows for testing if there is a difference in average daily log activity between vehicle and NTF treatment groups during the Infusion phase(s). This model assumes that the average activity for NTF and vehicle groups are the same during the MPTP phase adjusted for baseline activity and weight (a test was performed and found this difference to be insignificant).

$$Y_{ij} = \beta_0 + \beta_1(Baseline)_i + \beta_2(Weight)_i + \beta_3 I(t=1) + \beta_4 I(k=1) * I(t=1) \quad (4.1)$$

$$i = \text{Monkey} \quad , \quad j = \text{Day}$$

$$t = \begin{cases} 0 & \text{MPTP Period} \\ 1 & \text{NTF Infusion period} \end{cases} \quad , \quad k = \begin{cases} 0 & \text{NTF Treatment} \\ 1 & \text{Vehicle} \end{cases} \quad , \quad cor(Y_{im}, Y_{in}) = \rho, m \neq n$$

Although we do not show the regression output here, there is neither a significant difference between the treatment groups during the infusion period, nor is there a significant difference between the MPTP and treatment periods. However, the direction of the coefficients are consistent with the hypothesis that there is a treatment effect.

Given that these are invasive surgeries which take time to recover from, models were also fit looking at just the Infusion 3 period activity, activity 14 days after the surgical infusion procedure and a combination of these two conditions. Again, no significant differences were found between the time periods and the treatment groups. Figure 4.1 shows the daily average log(1+count/minute) in just the post-MPTP and infusion periods. For the infusion periods we only consider days more than 2 weeks (14 days) from surgery. Essentially this is the a subset of Figure 2.3 but at a higher resolution. The figure is consistent with the results of no treatment effect on average activity levels.

Figure 4.1: *Post-MPTP Vs Lagged Infusion Activity*

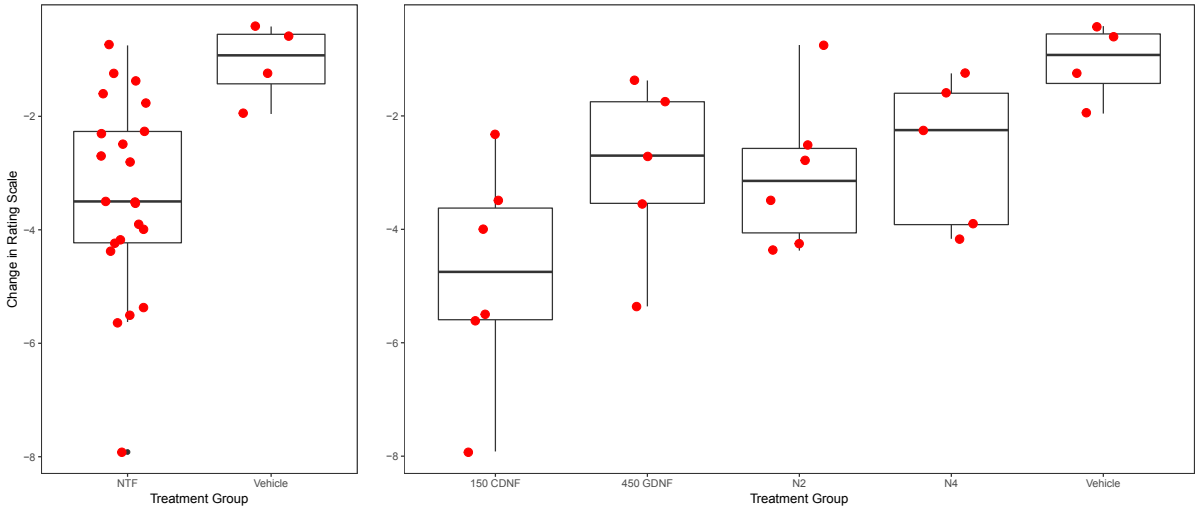
4.3 Change in Rating Scale

Considering the inability to detect a statistically significant difference in mean activity counts between the treatment/vehicle monkeys, it is of interest whether there is a measurable recovery in the monkeys receiving NTF that is not captured i mean activity. To that end, the change in Parkinson's rating scale was assessed using multiple linear regression. Equation 4.2 explicitly defines the model fit.

$$\Delta\text{Rating Scale}_i = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 \text{Weight}_i \tag{4.2}$$

Adjusting for weight, those monkeys in the 4 NTF treatment groups of interest have an expected 2.33 (95% CI: 0.46, 4.21) unit decrease in rating scale compared to the vehicle group. This difference indicates that the treatments are successful in reducing the Parkinson's symptoms associated with MPTP to a degree that is greater than no treatment. Figure 4.2 shows the distribution of change in rating scale. The left panel combines all the treatment groups under the label 'NTF' while the right panel displays the distributions for each NTF treatment group and the vehicle. It is clear that the 150 CDNF treatment group showed the largest improvements among all groups, while the three remaining NTF groups had approximately the same degree of improvement.

Figure 4.2: *Distribution of the Change in Rating Scale*



5. Conclusions

As wearable devices become smaller and less expensive, the use of actigraphy devices is likely to rapidly increase. Given the wealth of knowledge supporting the benefits of physical activity to human health, and the known deleterious effects of many chronic diseases on activity, it is imperative that statistical methods are developed to make inference from this complex data. Here, an attempt was made to show preliminary evidence that activity and/or patterns of activity can be used to elicit latent susceptibility to developing a debilitating disease (Parkinson's) and help understand aspects of potential recovery.

5.1 Response to MPTP

From the analysis described above, it can be concluded that indeed, MPTP does significantly affect average daily activity adversely in rhesus monkeys. However, the level of activity prior to MPTP administration is not predictive of response to MPTP as measured by the Parkinson's rating scale. There is moderate support in the data for the belief that daily patterns of activity can predict response to MPTP. While this support does not manifest in statistical significance, the association between 'activity patterns' and response to MPTP is stronger than the association with average levels of activity. It is unclear if this lack of statistical significance is due to a lack of power or a true lack of signal. Considering the relatively strong predictive power among monkeys that respond negatively to MPTP (i.e. rating scale ≥ 2), more work needs to be done to determine if it is possible to predict which monkeys will have a meaningful response to MPTP.

Then, among those that will respond negatively to MPTP, what measures can predict the severity of this effect.

5.2 NTF Infusion Effect

A significant improvement in Parkinson's symptoms was seen as measured using the Parkinson's rating scale for at least some of the neurotrophic factors measured. This improvement in symptoms does not appear to translate into a recovery in average activity. This could suggest that perhaps the neurotrophic factors would show improved outcomes in activity given a longer period of observation. Perhaps while the vehicle monkeys would continue to decline in activity, the neuroprotective effects of the treatments might allow monkeys to maintain their post-MPTP activity levels for a longer period than they would have been able to otherwise.

6. Bibliography

Auguie, B. (2012). gridExtra: functions in Grid graphics. R package version 0.9.1. http://CRAN.R-project.org/package=gridExtra

J. Bai, B. He, H. Shou, V. Zipunnikov, T. A. Glass, C. M. Crainiceanu. Normalization and extraction of interpretable metrics from raw accelerometry data, Biostatistics, 2014 Jan; Volume 15 (1): 102-116.

Burns, R. (1983). A Primate Model Of Parkinsonism: Selective Destruction Of Dopaminergic Neurons In The Pars Compacta Of The Substantia Nigra By N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. Proceedings of the National Academy of Sciences, 4546-4550.

Chassain, C., Eschalier, A., & Durif, F. (2001). Assessment of motor behavior using a video system and a clinical rating scale in parkinsonian monkeys lesioned by MPTP. Journal of Neuroscience Methods, 111(1), 9-16.

Crainiceanu, C., Reiss, P., Goldsmith, J., Huang, L., Huo, L., Scheipl, F. (2013). refund: Regression with Functional Data. R package version 0.1-9. http://CRAN.R-project.org/package=refund

Dahl, B. (2014). xtable: Export tables to LaTeX or HTML. R package version 1.7-3. http://CRAN.R-project.org/package=xtable

Nychka, D., Furrer, R., Sain, S. (2013). fields: Tools for spatial data. R package version 6.9.1. http://CRAN.R-project.org/package=fields

Dunn, O.J. Multiple Comparisons Among Means. Journal of the Americal Statistical Association, 1961; Volume 56 (293):52-64.

Dutta, S., Singh, G., Sreejith, S., Mamidi, K., Husin, J., Datta, I., Pal, R., Das, A. (2013). Cell Therapy: The Final Frontier for Treatment of Neurological Diseases. CNS Neuroscience & Therapeutics, 19(1), 5-11. Retrieved March 22, 2015.

Fitzmaurice, G., & Laird, N. (2004). Applied longitudinal analysis. Hoboken, N.J.: Wiley-Interscience.

Gomez-Mancilla, B., Boucher, R., Gagnon, C., Paolo, T., Markstein, R., & BeDard, P. (1993). Effect of adding the D1 agonist CY 208-243 to chronic bromocriptine treatment. I: Evaluation of motor parameters in relation to striatal catecholamine content and dopamine receptors. Movement Disorders, 8(2), 144-150. Retrieved March 22, 2015.

Hastie, T., & Tibshirani, R. (1990). Generalized additive models. London: Chapman and Hall.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models, cubic splines and personalized likelihood. Toronto: University of Toronto, Dept. of Statistics.

Højsgaard, S., Halekoh, U. & Yan J. (2006) The R Package geepack for Generalized Estimating Equations Journal of Statistical Software, 15, 2, pp1–11

Kong, P., Zhang, B., Lei, P., Kong, X., Zhang, S., Li, D., Zhang, Y. (2015). Neuroprotection of MAO-B inhibitor and dopamine agonist in Parkinson disease. International Journal of Clinical and Experimental Medicine, 431-439.

Liang, K.Y. & Zeger, S. Longitudinal Data Analysis Using Generalized Linear Models: Biometrika, 1986; Volume 73 (1): 13-22.

Mclean, M., Hooker, G., Staicu, A., Scheipl, F., & Ruppert, D. (2012). Functional Generalized Additive Models. Journal of Computational and Graphical Statistics.

Ramsay, J., Wickham, H., Graves, S., Hooker, G. (2013). fda: Functional Data Analysis. R package version 2.4.0. http://CRAN.R-project.org/package=fda

Ramsay, J., & Silverman, B. (2005). Functional data analysis (2nd ed.). New York: Springer.

Ruppert, D. (2002). Selecting the Number of Knots for Penalized Splines. Journal of Computational and Graphical Statistics, 735-757.

Schrack, J.A., Zipunnikov, V., Goldsmith, J., Bai, J., Simonsick, E.M., Crainiceanu, C.M., Ferrucci, L. Assessing the "Physical Cliff:" Detailed quantification of age-related differences in daily patterns of physical activity, Journal of Gerontology: Medical Sciences, 2014; Volume 69 (8): 973-979.

Stephen W., Spiro J. R. (2001). Comparing different methodologies used in wrist actigraphy. Sleep Review. :4042.

Wickham, H. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

Wickham, H. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12),1-20. URL http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36.

Wood, S.N. (2006) Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Xiao, L. (2013). fbps: sandwich smoother and fast covariance estimation. R package version 1.0.

Yan, J. & Fine, J.P. (2004) Estimating Equations for Association Structures Statistics in Medicine, 23, pp859–880.

Yan, J (2002) geepack: Yet Another Package for Generalized Estimating Equations R-News, 2/3, pp12-14.

Yue, X., Hariri, D.J., Zhang, S., Bartlett, M.J., Kaut, O., Mount, D.W., Wüllner, U., Sherman, S.J., Falk, T. (2014). Comparative study of the neurotrophic effects elicited by VEGF-B and GDNF in preclinical in vivo models of Parkinson's disease. Neuroscience, 258, 385-400. Retrieved March 22, 2015.